
Does Activating Sick-Listed Workers Work?

Evidence from a Randomized Experiment

Kai Rehwald*
Michael Rosholm*
Bénédicte Rouland**

2015/25

(*) Aarhus University
(**) LEMNA, Université de Nantes

Does Activating Sick-Listed Workers Work? Evidence from a Randomized Experiment

Kai REHWALD* Michael ROSHOLM† Bénédicte ROULAND‡§

This version: November 2015

Abstract

Using data from a large-scale randomized controlled trial conducted in Danish job centers, this paper investigates the effects of an intensification of mandatory return-to-work activities on the subsequent labor market outcomes for sick-listed workers. Using variations in local treatment strategies, both between job centers and between randomly assigned treatment and control groups within a given job center, we compare the relative effectiveness of alternative interventions. Our results show that the use of partial sick leave increases the length of time spent in regular employment and non-reliance on benefits, and also reduces the time spent in unemployment. Traditional active labor market programs and the use of paramedical care appear to have no effect at all, or even an adverse effect.

JEL Classification: J68, C93, I18

Keywords: Long-term Sickness; Vocational Rehabilitation; Treatment Effects; Randomized Controlled Trial

1 Introduction

As highlighted in the OECD (2010) report on *Sickness, Disability and Work*, sickness policy is rapidly moving to center stage in the economic policy agenda of most OECD countries. Budgetary considerations are one of the key reasons for this. Expenditure on paid sick leave in OECD

*K. Rehwald (krehwald@econ.au.dk): Department of Economics and Business Economics, Aarhus University

†M. Rosholm (rom@econ.au.dk): Department of Economics and Business Economics, Aarhus University

‡B. Rouland (Benedicte.Rouland@univ-nantes.fr): LEMNA-TEPP, University of Nantes. Correspondence to: Bénédicte Rouland, Université de Nantes IEMN-IAE, Chemin de la Censive du Tertre, BP 52231, 44322 Nantes Cedex 3, France.

§We would like to thank Knut Røed as well as seminar participants at CREST (Paris), IFAU (Uppsala), the University of Nantes, IZA (Bonn) and SFI (Copenhagen), and participants at the 2013 TEPP Conference “Research in Health and Labour”, the 2013 DGPE Workshop, the 2013 Annual Workshop of the Centre for Research in Active Labour Market Policy Effects (CAFÉ), the 2014 Annual Meeting of the Society of Labor Economists (SOLE), the 9th Nordic Summer Institute in Labour Economics, the “Empirical Strategies” Mini-Workshop in Bergen, the 2014 Annual Conference of the European Economic Association (EEA) and the 2015 Journées Louis-André Gérard-Varet in Public Economics, for their helpful comments and discussions. We gratefully acknowledge the financial and material support provided by the Danish Labor Market Authority. All remaining errors and shortcomings of the article are our own.

countries amounted on average to 0.8 percent of GDP in 2007.¹ Although this figure might seem rather low, it is nevertheless a matter of great concern in the current context of growing public deficits and debt burdens. In comparison, public spending on unemployment benefits reached “only” 0.55 percent of GDP in the same year.² Furthermore, absence due to sickness also implies a reduced labor supply, lost production, and health-related costs.³

Beyond the financial aspect of paid sick leave, the reintegration of sick-listed workers into the labor market is also a matter of great concern. Empirical research on the labor market has shown that frequent and/or long-term spells of absence are associated with a higher risk of unemployment (Hesselius, 2007), and can significantly reduce a worker’s subsequent earnings or prospect of employment (Markussen, 2012). The probability that a worker will then become inactive and dependent on a permanent disability pension also increases.

The importance of well conceived sickness policies is clear in this context. Such policies are essential both for the sick-listed worker (in terms of his/her successful reintegration) and for society as a whole (in view of budgetary constraints). Sickness policies have recently shifted from being passive towards a more employment-orientated approach, aiming both at reducing benefit dependency and increasing rates of employment.⁴ Taking Denmark as an example, vocational rehabilitation measures were implemented in 16 percent of all periods of sickness benefit in the years 2009-2011, compared to only seven percent in the period 2005-2007 (Boll et al., 2010). Vocational rehabilitation includes traditional active labor market programs (e.g., internships), paramedical care (e.g., physical therapy), and graded return-to-work (partial sick leave).

Our aim in this paper is twofold. First, we wish to assess the effects of an intensification of return-to-work activities on sick-listed workers’ subsequent labor market outcomes. Second, we aim to compare the relative effectiveness of the alternative interventions. Specifically, we use results from a large-scale randomized controlled experiment conducted in Danish job centers in 2009 among newly registered sick-listed workers. The treatment lasted four months and consisted of a combination of weekly meetings with caseworkers and intensive mandatory return-to-work activities in the form of graded return-to-work (partial sick leave), traditional activation, and/or paramedical care.

Our empirical strategy and key results can be summarized as follows. We first rely on a simple difference-in-means approach to identify the causal effect of offering a more intensive treatment package on the subsequent labor market outcomes for newly sick-listed workers. Specifically,

¹OECD data on social expenditure, taken from the OECD (2010) report on *Sickness, Disability and Work*. The term ‘sickness’ refers to public and mandatory private paid sick leave programs (occupational injury and other sickness-related daily allowances).

²OECD data on Labour Market Programmes, extracted from OECD data bank (<http://stats.oecd.org/>). As for Denmark—the country under consideration in this paper—expenditure on paid sick leave amounted to 1.4 percent, while public spending on unemployment benefit reached 0.96 percent of GDP in 2007.

³According to the Danish Ministry of Employment, absence due to sickness (short and long term) in 2006 reduced the supply of labor by five percent, which implies a cost of more than two percent of GDP.

⁴Refer to OECD (2010) for an outline of the main trends in recent reforms across the OECD.

we estimate a causal *intention-to-treat* effect on four outcome variables: accumulated weeks in regular employment, self-sufficiency (i.e., all forms of non-reliance on benefits), sickness, and unemployment. Second, in the spirit of Markussen and Røed (2014), we exploit variation in local treatment strategies, both between job centers and between treatment and control groups within a given job center, to compare the relative effectiveness of the alternative measures used.

Our findings reveal firstly that the experimental intervention as a whole has been ineffective. Sick-listed workers initially assigned to the treatment group spent less time in regular employment and self-sufficiency than their peers in the control group. Nevertheless, our results also show that a greater emphasis on offering graded return-to-work programs is associated with an increase in regular employment and self-sufficiency, and lower unemployment. On the other hand, traditional activation and paramedical care appear to have either no impact at all, or even an adverse impact. Taken in the round, our results suggest that programs focusing on graded return-to-work are the most effective in improving sick-listed workers' subsequent labor outcomes. These programs are associated with strong and long-lasting effects, but only for workers sick-listed from regular employment and for those with physical (non-mental) disorders.

In line with the rich literature on the effectiveness of active labor market policies for unemployed workers (see Card et al. (2010) for a meta-analysis), our study relates to the expanding literature on the impacts of return-to-work policies for (long-term) sick-listed workers.⁵ Return-to-work can be associated with various forms of interventions, including workplace-based⁶, educational, medical, and social interventions. The results are mixed, however; Frölich et al. (2004) for example, found that rehabilitation programs for the long-term sick (more than four weeks) in Sweden had no favorable effects at all, but that workplace interventions were less damaging than the alternative strategies. In a randomized study of the inflow of Swedish sick-listed individuals, Engström et al. (2015) found some negative effects associated with having early meetings to assess individuals' work capacity (more sickness absence and a higher probability of receiving disability benefits). In contrast, Everhardt and de Jong (2011) found strong positive impacts of return-to-work activities for long-term (nine months) sick employees in the Netherlands in terms of their likelihood of returning to work.

Some of the literature on workplace-based interventions focuses specifically on the effects of graded return-to work programs, i.e., some combination of part-time work and sickness benefits.⁷

⁵There is also another branch of the literature that relates to the impacts of return-to-work policies for temporary disabled workers. See for instance Aakvik et al. (2005) and Markussen and Røed (2014) for a study of the Norwegian Vocational Rehabilitation program. While there is no absolute definition of long-term sick leave, workers typically call on a temporary disability insurance system following a period of sick pay (which is more generous than the disability insurance); however this is available only for a limited period of time (usually around one year).

⁶Reviewing recent medical research, Van Oostrom et al. (2009) concluded that workplace interventions are effective in reducing sickness absence among workers with musculoskeletal disorders compared with normal forms of care, although they are not effective in improving health outcomes.

⁷Partial sick leave and partial sickness benefits are currently available in Sweden, Norway, Denmark, and

While accurate and reliable evidence remains scarce, Markussen et al. (2012) provides an exception. Using data collected from Norwegian administrative registers, the authors concluded that the use of graded (partial) rather than non-graded (full) sickness absence certificates reduces the length of periods of absence, and significantly improves the propensity for employment in subsequent years. Andrén and Svensson (2012) found that Swedish employees with musculoskeletal disorders assigned to part-time sick leave were more likely to recover to full work capacity than those assigned to full-time sick leave. From a randomized controlled trial performed in Finland among 63 patients with musculoskeletal disorders, Viikari-Juntura et al. (2012) showed that part-time sick leave reduced both the time taken to return to regular duties and the amount of sickness absence in the one-year follow-up period. In the Danish graded return-to-work program, Høgelund et al. (2010) found that participation in such a program significantly increased the probability that sick-listed workers returned to regular working hours. However, Nielsen et al. (2014) showed that its effect on the return to self-support differed substantially among the municipalities, and therefore warned against generalizing the results of the study to other Danish municipalities. Moreover, Høgelund et al. (2012) found no impact of the Danish graded return-to-work program for workers with mental health problems.

Based on a large-scale experimental design, the present study adds to the existing literature by offering a comprehensive evaluation of intensive mandatory return-to-work activities (activation requirements). In particular, we focus not only on workplace-based interventions but also on paramedical care, and thus compare the relative effectiveness of alternative intensive interventions (traditional activation vs. paramedical care vs. partial sick leave). We also consider all kind of diseases rather than focusing on a specific subsample of sick-listed workers as others have done.

The rest of the paper is organized as follows. Section 2 includes details of the randomized experiment. Section 3 is a description of the data and variables. Section 4 provides an explanation of the empirical strategy, and section 5 contains the findings. Some conclusions are given in Section 6.

2 The randomized experiment

The Danish sick-leave policy In Denmark, all employees, all self-employed, and all individuals receiving unemployment insurance benefits are entitled to receive compensation for each day they cannot work due to sickness (whether the sickness is work-related or not), provided they have worked at least 120 hours within the thirteen successive weeks prior to their sickness absence. Although the benefit period can be extended to more than a year under certain spe-

Finland. The authorities have strongly promoted the use of these in recommending partial sick leave as the mechanism of choice, where sick leave is needed. See Kausto et al. (2008) for a review of the use of partial sick leaves in the Nordic countries. A similar arrangement has also been in place in the UK since 2010 (known here as the “Fit Note”).

cific conditions, sickness benefits are normally available for a maximum of 52 weeks within an eighteen-month period. The employer finances the first 21 days of the sickness absence⁸, while municipalities are then responsible for funding the remaining period.

The municipalities play a key role in the return-to-work process. Besides managing sickness absence and work rehabilitation, it is their responsibility to monitor and assess recipients of sickness benefits. More precisely, municipalities must conduct an assessment of all sickness benefit cases no later than the eighth week of the sickness absence, and every fourth week from then on (or every eighth week in less complicated cases). Follow-up assessments must rely on updated and coordinated medical, social, and vocational information. The aim of these mandatory follow-up interviews is first to verify that the sick-listed individual is actually eligible for the benefit (i.e., s(he) has a work incapacity) and second, to help him or her to return to work as quickly as possible. Municipal case managers can implement various vocational rehabilitation measures, from job counselling to wage-subsidized job training and professional courses, and including graded return-to-work (partial sick leave).

It is worth noting that municipalities have economic incentives to reintroduce sick-listed individuals to the workplace because the state reimburses their expenditure on sickness benefit to varying degrees, depending on whether any return-to-work activities are implemented or not. Municipalities also have an incentive both to reduce expenditure on sickness benefit because the entitlement to reimbursement only applies to cases lasting less than 52 weeks, and to use part-time (rather than full-time) sick-listing, because this also reduces the final burden on them.

Finally, if despite medical treatment and vocational rehabilitation the sick-listed worker is unable to return to ordinary employment, the municipality may refer him or her to a permanent wage-subsidized job (*fleksjob*) with reduced working hours and special tasks. To be eligible for a *fleksjob*, the sick-listed worker must have a permanently reduced work capacity of at least 50 per cent and be no older than 65. The main difference between a *fleksjob* and graded return-to-work is that subsidized employment in a *fleksjob* is granted for an unlimited time while participation in a graded return-to-work program is always temporary. If the sick-listed individual cannot return to a *fleksjob*, the municipality may award disability benefits.

The experiment In early 2009, the National Labor Market Authority in Denmark launched a randomized controlled trial (hereafter RCT) to test at a small scale some of the elements that were to be included in the forthcoming legislation on active interventions relating to those receiving sickness benefits.⁹ The overall purpose of this experiment was to examine whether sick-listed workers who behave in a more proactive way during their sickness period can achieve a

⁸At the time of the experiment (2009), the employer was required to finance the first fourteen days of the illness period; the rules on this changed in October 2009.

⁹A new and intensified treatment package was eventually rolled out nationally (with adjustments) in late 2009, when a new law governing the return-to-work of sick-listed workers was passed in Denmark.

greater degree of autonomy (in terms of returning to work and staying in work) than they would have done had they not behaved proactively. The present study is a report on the outcomes of this experiment.

The experiment was designed as an RCT conducted in 16 job centers across Denmark, with “random” assignment to treatment by birth year, even or odd. A total of 5,652 newly sick-listed workers were covered by the experiment, of which 2,795 were assigned to the control group and 2,857 to the treatment group.¹⁰ Individuals were notified that they had been assigned to the treatment group during the first follow-up meeting (no later than the eighth week of their sickness absence) by means of a standard letter from The National Labor Market Authority. Newly registered sick-listed workers born in odd years were subject to intensive efforts (treatment group), while those born in even years were subject to normal levels of effort (control group). Although assignment to treatment was predetermined, not random, we nevertheless consider our study to be an RCT; individuals were not made aware of the assignment mechanism, and a person’s birth year is “random” from his/her own perspective.

The treatment lasted 18 weeks and consisted of a combination of weekly meetings with caseworkers AND intensive mandatory return-to-work activities in the form of either a graded return-to-work (partial sick leave) and/or traditional activation and/or paramedical care. More precisely, *traditional activation* includes vocational guidance advice and courses aimed at enhancing skills, together with internships and on-the-job training. *Paramedical care* consists of courses on handling one’s own situation, psychological consultations, nutritional counselling, and exercise including back exercises and other physical training. Lastly, the aim of the *graded return-to-work* measure is to support employees with reduced work ability to continue and return to work via partial sick leave. This involves working part-time (a reduction in working hours by at least four hours per week) and receiving a partial sickness benefit for the hours off work, on top of a partial salary (at a normal hourly rate). The underlying idea is that most sick leave days are a result of non-communicable diseases, and a person’s work capacity while sick may be reduced but it is not nothing. The return to regular working hours should take place as soon as possible, and certainly within the 52-weeks payment period of full sickness benefits. Graded return-to-work must be implemented with the agreement of the employer, the sick-listed worker, and the municipality. In practice, either the sick-listed worker and the employer arrange a graded return-to-work on their own initiative and then ask the municipality to approve this, or the municipality determines that the sick-listed worker is able to work part-time and therefore asks him or her to agree a graded return-to-work with the employer. If the sick-listed worker refuses to enrol in the program even though the municipal case manager recommends it, the sick benefit can be reduced.

¹⁰Table A1 in the Appendix shows the number of individuals by employment region, job center and treatment status.

Within four weeks of the first interview, individuals in the treatment group were required to participate in some kind of program (graded return-to-work, traditional activation and/or paramedical care) for at least ten hours a week for up to four months. Table 1 provides details of the extent of these return-to-work activities. Compared with the control group, it is clear that the treatment group received more intensive programs of treatment. For treated individuals, the number of return-to-work activities (all three types) was higher, and the interventions generally began earlier and lasted longer.

Table 1: Number of return-to-work activities by type and treatment status

Type	Control group (N=2,795)	Treatment group (N=2,857)
Panel A: Traditional activation		
Total number of weeks in traditional activation	1,111	6,129
Total number of traditional activation spells	97	512
... of which counselling and training	81	417
... of which on-the-job training	16	95
Avg. week in which the 1st traditional activation spell begins ^a	6.5	4.6
Avg. length of traditional activation spells (in weeks) ^b	11.9	12.5
Panel B: Paramedical care		
Total number of weeks in paramedical care	505	11,314
Total number of paramedical care spells	42	1,107
... of which courses in handling one's own situation	19	228
... of which psychologist	1	64
... of which nutritional counselling	0	63
... of which exercise	16	326
... of which back exercises or other physical training	4	166
... of which other	2	260
Avg. week in which the 1st paramedical care spell begins ^a	5.4	3.5
Avg. length of paramedical care spells (in weeks) ^b	12.6	14.4
Panel C: Graded return-to-work		
Number of initiated programs	366	566
Avg. week in which the program begins ^a	5.7	4.5

^a Measured in weeks since first follow-up meeting with the municipal caseworker (typically eight weeks after the onset of illness).

^b When calculating the average length of activation spells in weeks, uninterrupted sequences of alternative activities are treated as a single spell (93 spells in the control and 492 spells in the treatment group). Similarly for the average length of paramedical care spells in weeks (40 spells in the control and 787 spells in the treatment group).

Moreover, the outcome from the first follow-up meeting with the municipal caseworker (eight weeks after the onset of illness) could also be a total program exemption for some sick-listed workers. As shown in Table 2, around 13 percent of all participants in the experiment were excused from any kind of return-to-work activity, mainly because therapy was a hindrance (52 percent of all cases of exemption), or because of a pregnancy-related illness (44 percent of cases). Exemption rates were higher in the treatment group (around 17 percent) than in the control group (almost ten percent). However, the exempted are still included in the analysis. Specifically, those exempted from the treatment group received the treatment as usual (i.e., the same as individuals in the control group) and are included in the treatment group. Those exempted from the control group also received the normal amount of return-to-work activity (even though they

were excused). We account for this significant number of “no-shows” by estimating intention-to-treat effects, so our results could be valid even given imperfect compliance.

Table 2: Number of individuals excused/not excused by reason and treatment status

	Control group	Treatment group
Excused	272	496
... of which early retirement	7	14
... of which terminal illness	3	2
... of which pregnancy-related illness	167	174
... of which therapy was a hindrance	95	306
Not excused	2,523	2,361
Total	2,795	2,857

Notes: The table shows the number of individuals excused/not excused by treatment status. The exempted are included in the analysis because exemption rates were higher in the treatment group than in the control group. Specifically, those exempted from the treatment group received the treatment as usual (i.e., the same as individuals in the control group) and are included in the treatment group. Those exempted from the control group also received the normal amount of return-to-work activity (even though they were excused).

Finally, it is worth noting that the job centers were responsible for the organization of the experiment. They carried out interviews and decided on the composition and content of any vocational rehabilitation, accounting for the individuals’ needs and adapting to local conditions. Therefore, some variations in treatment intensity between job centers may be assumed. From Figure 1, it is very clear that there is substantial variation in the intensity of return-to-work activities, both between the 16 job centers covered by the experiment, and between treatment and control groups within a given job center.¹¹ As we explain in Section 4, we exploit this variation in local treatment strategies to compare the effectiveness of the alternative return-to-work activities (graded return-to-work, traditional activation, and paramedical care).

3 Data and variables

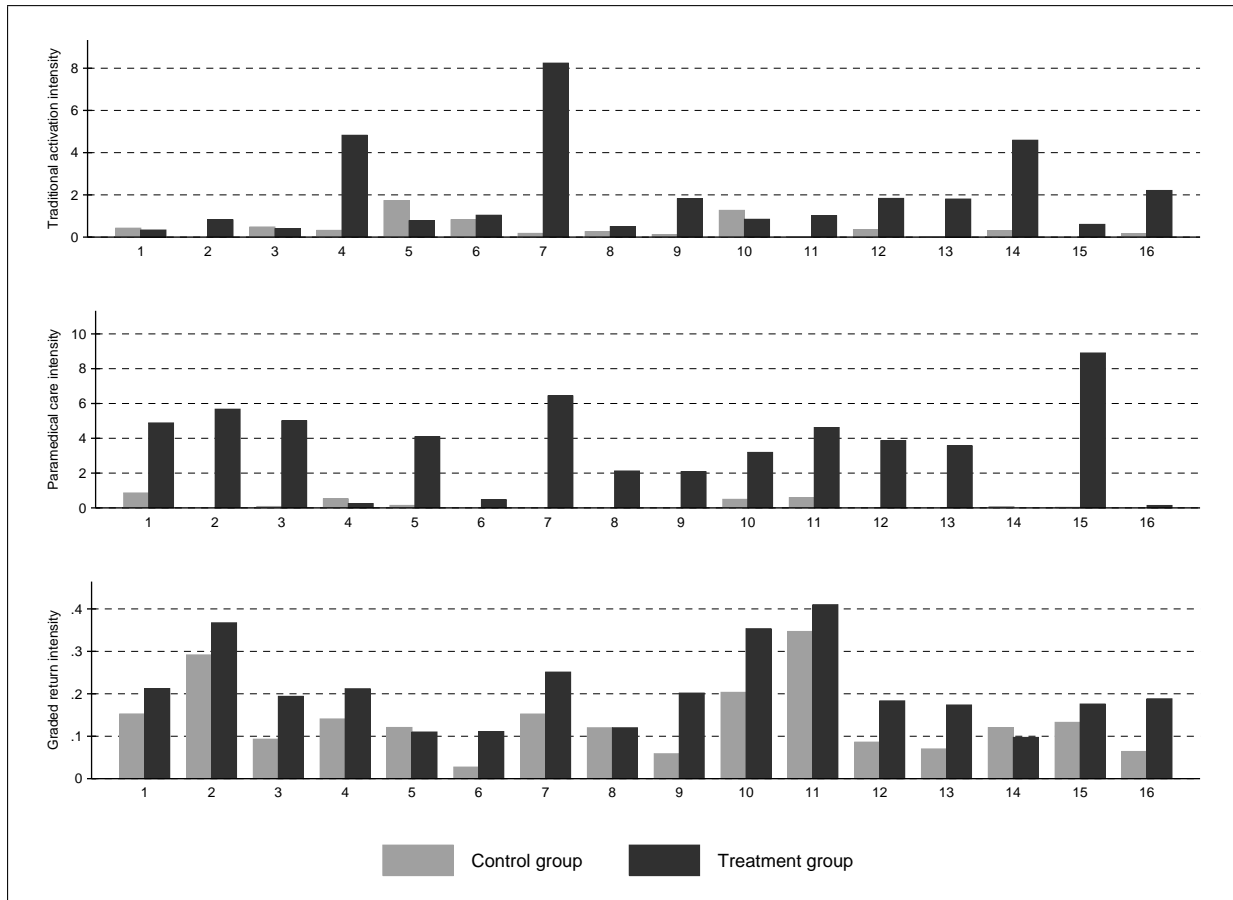
The empirical analysis is based on four different data sets. First, we exploit unique Danish data derived from the controlled field experiment described above. These data include binary variables for each type of return-to-work activity (graded return, traditional activation and paramedical care) and for each meeting scheduled/held, as well as the number of hours per week in each type of activation and the timing of the intake. We can therefore follow participation accurately on a weekly basis. The data also include information about possible exemption from activation and the job center is identified in each case.

Second, from the DREAM register we obtained information about the type of social welfare benefit received in a given week¹², and about individual characteristics. DREAM is an amalga-

¹¹Similarly, the overall management of the usual vocational rehabilitation is regulated by law, but the assessment and the implementation of individual cases are controlled at job center level. Therefore, there is also substantial variation in the intensity of return-to-work activities among the control group.

¹²The classification scheme includes all types of public income transfer schemes in Denmark. If a citizen does

Figure 1: Traditional activation, paramedical care and graded return-to-work intensities by job center and treatment status



Notes: Indices 1 to 16 on the horizontal axes refer to job centers: (1) Bornholm, (2) Gentofte, (3) Greve, (4) København, (5) Ringsted, (6) Vordingborg, (7) Aalborg, (8) Morsø, (9) Randers, (10) Holstebro, (11) Herning, (12) Horsens, (13) Svendborg, (14) Nyborg, (15) Odense and (16) Aabenraa. Traditional activation intensities are calculated as the average number of traditional activation weeks per individual in the first 20 weeks after enrolment. Similarly for paramedical care intensities. Graded return intensities are calculated as the proportion of individuals participating in a graded return-to-work program within the first 20 weeks.

mation of information from several different sources, and is updated once per month by the The National Labor Market Authority.¹³ Data from the DREAM register were obtained from the week in which the experiment began (i.e., between the first week of January 2009 and the third week of November 2009) up to the end of 2012, allowing a three-year follow-up. Specifically, we constructed four outcome variables from the DREAM database, using the weekly information on social transfer payments, namely the cumulative number of weeks spent in regular employment (i.e., employed), in non-reliance on benefits (i.e., self-sufficiency), in sickness and in unemployment, where *self-sufficiency* covers the officially designated success criteria and covers all periods of non-benefit (self-sufficiency without employment, employment, and ordinary education).¹⁴

Besides the treatment package, our explanatory variables include age, gender, marital status, country of origin (three categories—Denmark, western, and non-western countries), the state before sickness (three categories—job, unemployment benefits and self-sufficiency), the duration of elapsed sickness at the start of the experiment, the proportion of time spent on sickness payments (in each of the three years before the sickness), and the same proportion for time spent on public income support of any kind (also in each of the three years before the sickness). Table 3, which shows means and standard deviations of the explanatory variables by treatment status, suggests that there are very few observed differences between the treatment and control groups.¹⁵ “Random” assignment based on birth year—even or odd—successfully balanced groups. The most significant difference, though nonetheless, relates to the status before sickness: the treatment group was on average in regular employment slightly less than the control group (76.3 vs. 79.9 percent), and in unemployment slightly more (17.4 against 14.2 percent).

Third, from Statistics Denmark we obtain data on socio-economic characteristics at the municipality level. In particular, for each municipality, we collected annual data on the total fertility rate, average age, and life expectancy for new born babies. We obtained quarterly information on the proportion of the working-age population with no more than ten years of schooling, on external and internal migration as a percentage of the population, and on the number of reported criminal offences per capita. As for the local labor market conditions, we calculated quarterly ratios as a percentage of the labor force for the working-age population outside the labor force, or receiving sickness benefits, or employed in the primary sector, as well

not receive social benefits in a given period, it is represented by empty week-variables.

¹³The week-variables are only allowed to contain one type of compensation code at a time. This implies that the types of social benefits are ranked. The ranking implies that if a citizen changes the type of social benefit in the middle of a week, only the highest ranked type is registered that week.

¹⁴More precisely, we first generated indicators for regular employment, sickness, and unemployment using the variable “status” from the DREAM register (i.e., the weekly information on labor market status). The outcome variable *regular employment* is defined using status 500 (i.e., the regular employment indicator is equal to one if the variable status is equal to 500). Similarly, the sickness indicator is based on statuses 890 to 899, while the unemployment indicator is generated from statuses 111, 112, 124, 125, 130 to 138, 200 to 300, and 730 to 738. Having defined these indicators, we computed the cumulative number of weeks spent in each state.

¹⁵Table A2 in the Appendix shows the means and standard deviations of the explanatory variables by treatment status for the sample of 4,728 individuals (see later).

Table 3: Pre-treatment characteristics by treatment status (N=5,652)

Variable	Control group		Treatment group		Difference
	Mean	Std. dev.	Mean	Std. dev.	
Panel A: Socio-demographic characteristics					
< 30 years old	0.161	(0.367)	0.144	(0.351)	-0.016*
30-39 years old	0.244	(0.429)	0.238	(0.426)	-0.006
40-49 years old	0.259	(0.438)	0.269	(0.444)	0.010
> 49 years old	0.337	(0.473)	0.349	(0.477)	0.012
Male	0.407	(0.491)	0.427	(0.495)	0.021
Married	0.597	(0.491)	0.581	(0.493)	-0.016
Danish origin	0.875	(0.330)	0.877	(0.328)	0.002
Western origin	0.046	(0.210)	0.047	(0.211)	0.001
Non-western origin	0.078	(0.269)	0.076	(0.264)	-0.003
Panel B: Labor market history					
Sick-listed from regular empl.	0.799	(0.401)	0.763	(0.426)	-0.036***
Sick-listed from UI-benefits	0.142	(0.350)	0.174	(0.379)	0.031***
Sick-listed from self-empl.	0.059	(0.236)	0.063	(0.244)	0.004
Elapsed sickness duration ^a	7.841	(6.414)	7.621	(6.258)	-0.220
Time spent on sickness-ben. ^b	0.188	(0.147)	0.188	(0.149)	-0.001
Time spent on sickness-ben. ^c	0.075	(0.174)	0.076	(0.173)	0.001
Time spent on sickness-ben. ^d	0.066	(0.164)	0.064	(0.162)	-0.003
Degree of pub. inc. support ^{b e}	0.302	(0.257)	0.311	(0.263)	0.009
Degree of pub. inc. support ^{c e}	0.220	(0.320)	0.220	(0.321)	-0.001
Degree of pub. inc. support ^{d e}	0.244	(0.343)	0.243	(0.340)	-0.001
Panel C: Diagnoses					
Back and neck disorders	0.142	(0.367)	0.160	(0.367)	0.018*
Moving incapacity	0.189	(0.391)	0.195	(0.396)	0.006
Musculoskeletal disorders	0.035	(0.184)	0.046	(0.209)	0.011**
Cardiovascular diseases	0.048	(0.214)	0.047	(0.212)	-0.001
Stomach/liver/kidney diseases	0.036	(0.188)	0.031	(0.174)	-0.005
Mental health problems	0.300	(0.458)	0.292	(0.455)	-0.008
Chronic pain	0.007	(0.082)	0.009	(0.093)	0.002
Cancer	0.032	(0.176)	0.038	(0.191)	0.006
Other	0.211	(0.408)	0.183	(0.386)	-0.029***
Number of observations	2,795		2,857		

^a at start of experiment (in weeks) ^b in year before sickness ^c in second last year before sickness ^d in third last year before sickness ^e any kind of public income support.

Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

as the number of full-time unemployed each quarter as a percentage of the labor force.

Finally, *jobindex.dk* is a collection of all the vacancies posted on the internet (online newspapers, job centers, job databases, etc.), which provides us with monthly information on the number of open vacancies and newly opened vacancies per unemployed person. We used this information to control for local environment characteristics when exploiting the variation in treatment strategies across job centers.

4 Empirical strategy

4.1 The effect of the treatment bundle as a whole

Our first aim is to identify the overall causal effect of offering the intensified treatment package on sick-listed workers' subsequent labor market outcomes. We may write

$$Y_i = \beta_0 + \beta_Z Z_i + \mathbf{X}_i \boldsymbol{\beta} + \varepsilon_i \quad (1)$$

where Y_i is the outcome of individual i (we consider four different measures, see below), Z_i is a treatment status indicator equal to unity for clients assigned to the treatment group (zero otherwise), and \mathbf{X}_i is the vector of pre-treatment characteristics summarized in Table 3 (with age entering linearly, not as a dummy-coded categorical predictor). Because treatment assignment is essentially random, Z_i will, by design, be independent of \mathbf{X}_i and of any other (observed or unobserved) pre-treatment variable.¹⁶ The coefficient β_Z corresponds to the (conditional) difference in the means of outcome variable Y_i between treated and controls. It identifies the *intention-to-treat* (ITT) effect $E(Y_i|Z_i = 1) - E(Y_i|Z_i = 0)$ if $E(\varepsilon_i|Z_i) = E(\varepsilon_i)$. This is likely to hold given that Z_i is assigned at random. The experimental impact estimate $\hat{\beta}_Z^{OLS}$ can be interpreted causally; it is unbiased and consistent.

Regarding the dependent variable in equation (1), Y_i , we consider four measures evaluated at different points in time: the number of weeks in (i) regular employment, (ii) self-sufficiency, (iii) sickness, and (iv) unemployment during the first, second, and third year after randomization.¹⁷ The self-sufficiency measure is meant to cover all forms of non-benefit receipt. It encompasses individuals in regular (i.e., wage) employment, as well as self-employed, housewives and everyone else not receiving public income transfers.¹⁸

Additionally, in order to trace the trajectory of treatment effects over time in greater detail, let Y'_i be an alternative set of response variables denoting the cumulative total number of

¹⁶It immediately follows that conditioning on \mathbf{X}_i should leave $\hat{\beta}_Z^{OLS}$ unaffected. At the same time, we expect a reduction in the residual variance to be accompanied by an increase in precision.

¹⁷All individuals can be followed for 141 weeks, and outcome variables referring to the third year (up to 156 weeks) after enrolment are non-missing for all but eight individuals.

¹⁸Exceptions are (subsidized) adult apprentices ("Voksenl rlinge") and individuals receiving state educational support ("Statens Uddannelsesst tte"). Both groups are considered to be self-sufficient.

weeks—running sums counting from the week of intake—in each of the four labor market statuses described above. We evaluate each of these outcomes at each week starting from the week of (individual) intake and ending with the 156th week after randomization (implying a total of 156 regressions per outcome variable). The results are presented graphically.

Besides its ease of use, the main advantage of the identification strategy discussed above is that it produces informative estimates even in the presence of imperfect compliance. Recall that there are a significant number of “no-shows”: more than one out of every six sick-listed workers assigned to treatment received treatment as usual (see Table 2). Under these conditions, the proposed experimental evaluation design (ITT) is the most relevant from a policy perspective. In fact, one of the aims of the experiment was to test the intensified treatment package on a small scale before it was eventually rolled out nationally with minor adjustments (impact pilot).

The main drawback, however, is that a simple comparison of average labor market outcomes between treated and non-treated (“difference in means”) only allows us to evaluate the treatment package *as a whole*. Does the package work? While this question is relevant, it is clearly not the only causal relationship of interest. In particular, we would like to attribute the overall performance of the treatment bundle to its individual components (traditional activation, paramedical care, graded return-to-work). Yet, since the experiment was not designed as a multi-arm trial (the ideal experiment for answering this richer research question), isolating their individual impacts is a challenge. A naive comparison of those who did and those who did not participate in a particular treatment activity risks being biased by self-selection: observed differences in labor market outcomes are equal to the sum of the effect of the treatment on the treated and selection bias. Unbiased estimates can only be obtained if participation is as good as randomly assigned conditional on covariates (conditional independence)—a very strong assumption to make. In consequence, the “difference in means approach” does not enable us to compare the relative effectiveness of the three alternative policies. Our second identification strategy, which is in spirit of that used by Markussen and Røed (2014), addresses this methodological challenge.

4.2 Disentangling the bundle, isolating effects

By exploiting variations in local treatment strategies, Markussen and Røed (2014) analyze the effectiveness of four alternative vocational rehabilitation programs for a large sample of temporary disability benefit recipients in Norway. We follow their approach.

To begin with, recall that there is substantial variation in treatment intensities both between the 16 job centers involved in the experiment and between the treatment and control groups within a given job center (Figure 1). As such, there are 32 distinct “local treatment environments”, each characterized by idiosyncratic working methods and treatment priorities. Differences in these “treatment cultures” may be seen in, e.g., the choice and combination of treatment types,

the speed with which newly registered clients are exposed to them, or the length of initiated activation spells. The resulting “treatment portfolio” is shaped by each of these choices, which may, in sum, be referred to as a “treatment strategy”. Our aim is to proxy these treatment strategies by vectors of local treatment strategy characteristics (φ_i). These can, in turn, be used to identify the effects of alternative interventions.

The vectors of local treatment strategy characteristics (φ_i) are individual-specific and depend on the treatment histories of all *other* sick-listed workers exposed to the same local treatment regime (more on this below). Each vector contains three elements— φ_{Si} (S =traditional activation, paramedical care, graded return-to-work)—and is meant to describe “both the choice of (first) treatment, and the speed with which it is implemented” (Markussen and Røed, 2014: 15). We estimate the vector of local treatment strategy characteristics in the framework of a linear discrete transition rate model with competing risks. In particular, we consider exits from a single state (“*sick and untreated*”) to multiple destinations: participation either in traditional activation, in paramedical care, or in a graded return-to-work program. Although a sick-listed individual can be exposed to a combination of alternative interventions over the course of his/her rehabilitation, our focus lies in the choice of first treatment. Accordingly, the data pattern used for estimation is characterized by single spells, one for each individual; repeat spells are ignored.

The survival time data at hand is an interval-censored inflow sample with weekly observations. Now, let P_{Sijd} be a destination-specific censoring variable equal to unity (zero otherwise) if sick-listed individual i , registered in local treatment environment j , makes a transition into treatment S after having been untreated—and thus at risk of making the transition—for d weeks. Given these event indicators, we organize the dataset in the following way. First, because individuals who are not sick-listed in the week of intake (which is the case for 924 out of 5,652 experimental units) are not at risk of being activated, they are excluded from the analysis beforehand; 4,728 individuals remain. Next, starting with a panel in person-week format, for each sick-listed worker we remove observations after the first transition into one of the three alternative programs. Uncompleted spells are right-censored in the absence of an event within the first 20 weeks—recall that the treatment period is meant to last only 18 weeks—or if the sickness spell ends. In short, the resulting panel is unbalanced and contains, for each sick-listed client, one observation per week at risk of being activated for the first time.

Then, letting \mathbf{D} denote a vector of duration dummies (one for each week), \mathbf{X}_i the vector of individual pre-treatment characteristics summarized in Table A2 in the Appendix (where all non-binary variables, including age, enter in a quadratic fashion to allow for flexibility) and \mathbf{X}_{jd} a vector of municipality-level controls referring to local treatment environment j in week d after randomization (socio-demographic characteristics and local labor market conditions; see Table A3 in the Appendix for a complete list), we estimate—separately for each of the three alternative

treatments—the following linear probability model:

$$P_{Sijd} = \beta_0 + \mathbf{D}\boldsymbol{\lambda}_S + \mathbf{X}_i\boldsymbol{\theta}_S + \mathbf{X}_{jd}\boldsymbol{\vartheta}_S + u_{Sijd} \quad (2)$$

Markussen and Røed (2014: 16) argue that the residuals in this model have an appealing interpretation. Particularly, the sum of individual residuals,

$$\hat{u}_{Sij} = \sum_{d=1}^{D_{Si}} \hat{u}_{Sijd} \quad (3)$$

where D_{Si} corresponds to the number of weeks sick-listed worker i was at risk of making the transition into treatment S , can “be interpreted as the estimated covariate-adjusted transition propensity at the claimant level”. Right-censoring and duration dependence aside, \hat{u}_{Sij} is equal to the (weighted) number of “lacked” waiting weeks for transition into S compared to what one would expect given the observed pre-treatment characteristics of client i (\mathbf{X}_i) and the municipality-level socio-economic characteristics of local treatment environment j (\mathbf{X}_{jd}). For instance, $\hat{u}_{Sij} > 0$ indicates that the transition happened earlier than expected.

Now, recall that the vector of local treatment strategy characteristics is intended to describe both the choice of first treatment and the speed with which newly registered clients are exposed to it. Specifically, the local treatment strategy characteristics relevant for individual i (φ_{Si}) may be defined as the average covariate-adjusted transition propensity of all sick-listed workers other than i subject to the same local treatment regime (leave-out mean):

$$\varphi_{Si} = \frac{1}{n_j - 1} \sum_{k \in N_j^{-i}} \hat{u}_{Skj} \quad (4)$$

where N_j^{-i} denotes the set of individual i ’s peers in local treatment environment j and $n_j - 1$ is the cardinality of this set. Note that even though individual i ’s own covariate-adjusted transition propensity does not enter equation (4), φ_{Si} is *not* completely exogenous to individual i : individual i ’s characteristics (\mathbf{X}_i) and treatment history (summarized by P_{Sijd}) influence the parameter estimates, predicted values, and residuals in (2), therefore all covariate-adjusted transition propensities \hat{u}_{Sij} in (3) and thus φ_{Si} in (4). To overcome this problem, we exclude individual i when fitting the linear probability model specified in (2). Accordingly, we estimate 4,728 linear probability models per treatment type S , excluding one individual at a time and computing one datapoint (φ_{Si} for the excluded individual) per iteration.

Equipped with the proxies for local area practice styles, we specify the following outcome equation:

$$Y_i = \beta_0 + \boldsymbol{\varphi}_i\boldsymbol{\alpha} + \mathbf{X}_i\boldsymbol{\gamma} + \mathbf{X}_j\boldsymbol{\omega} + \varepsilon_i \quad (5)$$

where Y_i is again the outcome of individual i (we consider the same measures as in (1)), \mathbf{X}_i and \mathbf{X}_j are the same individual and municipality-level socio-economic characteristics as in (2) and $\boldsymbol{\varphi}_i$ is the vector of local treatment strategy characteristics with elements φ_{Si} (S =traditional activation, paramedical care, graded return-to-work) as defined in (4).¹⁹ The coefficients $\boldsymbol{\alpha}$ identify the impacts of marginal changes in local treatment strategies on subsequent labor market outcomes and can thus be interpreted as *intention-to-treat* effects. Note that this interpretation hinges only on the assumption that local treatment strategies are as good as randomly assigned conditional on the individual and municipality-level controls included in (5). Provided that there is no unaccounted for purposeful sorting of sick-listed workers into treatment environments, this assumption seems plausible.

Following Markussen and Røed (2014), we also specify a model in which program participation indicators enter directly as right hand side variables. For this purpose, consider the following specification:

$$Y_i = \lambda_0 + \mathbf{P}_i\boldsymbol{\mu} + \mathbf{X}_i\boldsymbol{\delta} + \mathbf{X}_j\boldsymbol{\zeta} + \varepsilon_i \quad (6)$$

where \mathbf{P}_i is a vector whose elements P_{Si} (S =traditional activation, paramedical care, graded return-to-work) are indicators of actual treatment receipt. In particular, let P_{Si} equal unity (zero otherwise) if individual i participated in treatment activity S at some time during the first 20 weeks after enrolment.

As stressed earlier, one potential problem with (6) is that individuals may, at least in part, self-select into their preferred program, rendering \mathbf{P}_i endogenous if selection is based on unobserved traits. The parameters in (6) cannot then be consistently estimated by Ordinary Least Squares (OLS). To overcome this problem, we instrument the potentially endogenous elements of the vector of actual treatment receipt (P_{Si}) by the elements of the vector of local treatment strategy characteristics (φ_{Si}). These are, arguably, valid instruments: the instruments are *as good as randomly assigned* (*independent* of potential outcomes), *relevant* (partially correlated with the endogenous treatment receipt indicators) and *exogenous* (uncorrelated with the unobservable error term of the structural model). Under instrument validity, and assuming that there are no defiers (*monotonicity*), the coefficients $\boldsymbol{\mu}$ identify local average treatment effects (LATEs), i.e., the average causal effects of actually participating in alternative programs for the compliant subpopulation (Imbens and Angrist, 1994).

We argued above that exposure to a particular local treatment strategy is as good as randomly

¹⁹Note that besides \mathbf{X}_i (see Table A2) and \mathbf{X}_j (see Table A3), we are controlling for a large number of *week of intake dummies* when estimating model (5). The same applies to the linear probability model specified in (2) and to equation (6) below. Also note that the vector of municipality-level characteristics entering the linear probability model in (2), \mathbf{X}_{jd} , consists of time-varying variables, whereas the vector \mathbf{X}_j in (5) is time-invariant (variables refer to the week of intake). Arguably, transition propensities in (2) depend on *current* conditions, whereas labor market outcomes in (5) depend on *initial* conditions.

assigned and thus independent of potential outcomes, provided that there is no unaccounted for purposeful sorting of sick-listed workers into treatment environments. Also note that *independence* suffices for a causal interpretation of reduced form estimates and that equation (5) actually corresponds to the reduced form of the instrumental variables model specified in (6). Next, it should also be clear from the preceding discussion that the proposed instruments are, by construction, strongly correlated with the potentially endogenous vector of actual treatment receipt. Besides, the *instrument relevance condition* can (and will) be tested.

Regarding the *exclusion restriction*, we need to maintain that the only channel through which local treatment strategies affect labor market outcomes is through their effect on program participation. In particular, we need to assume that the instruments are uncorrelated with unobserved determinants of the outcome. In what follows, we will argue that this assumption is plausible. For this purpose, it is instructional to think of the error term in equation (6) as being composed of two parts. The first part contains unobserved determinants of labor market success that are peculiar to the individual sick-listed worker, i.e., unobserved individual characteristics such as ability, motivation, or the loss in work capacity due to sickness (which is not fully captured by controlling for diagnoses). The second part comprises all remaining factors, i.e., unobservables that are non-specific to a particular client. This second component consists first and foremost of (potentially unobserved) local labor market conditions and other municipality-level influences.

Now, note that, first, the set of instruments relevant for individual i is *completely* exogenous to individual i in the sense that neither individual i 's characteristics nor individual i 's treatment history have *any* impact on the instruments. Individual characteristics (observed or unobserved) should therefore be orthogonal to the instrument. It remains to be argued that the same is true for the second part of the error term. We need to maintain that local treatment cultures are uncorrelated with unobserved local labor market conditions and other municipality-level unobservables determining the outcome. If we think of "treatment cultures" as being the result of the interplay between national statutory provisions and a "combination of individual judgment, guesswork, personal experience, and convenience" (Markussen and Røed, 2014: 6), i.e., if a treatment environment's treatment priorities are first and foremost determined by factors unrelated to current local conditions, this requirement is arguably satisfied. Given this line of reasoning, it should also be the case that *observable* municipality-level variables have no significant effect on observed treatment portfolios. To test for selection on observables, we regress the treatment intensity indicators plotted in Figure 1 on the vector of municipality-level characteristics, \mathbf{X}_j , as described in Table A3 in the Appendix. For traditional activation and paramedical care, none of the estimated coefficients is statistically significant at conventional levels. Moreover, F statistics suggest that they are also jointly insignificant. For graded return-to-work, only one out of 16 covariates ends up being statistically significant (at the ten percent level). We take

this as evidence in favor of the assumption that there is no selection based on (observable) municipality-level variables; this supports the exclusion restriction.

5 Results

5.1 Evaluating the treatment package as a whole (difference in means approach)

To begin with, we present the results of evaluating the treatment package as a whole by comparing average labor market outcomes of treated and non-treated individuals. Does offering intensified services improve sick-listed workers' labor market prospects compared with clients receiving treatment as usual? The answer is given in Table 4. Panel A displays unconditional intention-to-treat effects (pure differences in means) *calculated* by regressing the dependent variables in columns I to IV on a treatment status indicator. Conditional intention-to-treat effects, *estimated* by partialling out the impacts of pre-treatment variables (socio-demographic characteristics, individual labor market history and diagnoses), are shown in Panel B.²⁰

It is immediately apparent from Table 4 that the experimental intervention as a whole is ineffective, not to say harmful. Offering the combined treatment package—a bundle consisting of intensified traditional activation, paramedical care, and graded return-to-work programs—has, on average, adverse impacts on subsequent labor market prospects when evaluated against the counterfactual outcomes of sick-listed workers receiving standard services.

Regarding the outcome variables regular employment and self-sufficiency, estimated intention-to-treat effects are negative, moderate in size, and statistically significant (at the ten, five, or one percent level) during the first and second year after random assignment. The estimates suggest, for instance, that offering intensive rather than standard services reduces the time spent in self-sufficiency (non-benefit receipt) by one week on average during the first year, and one and a half weeks during the second year (Panel B). The corresponding estimates for regular employment are slightly smaller in absolute terms, but still sizeable, given that the non-treated spent on average only 16.5 weeks in regular employment during the first year and 21 weeks during the second year after randomization (Panel A). The impacts during the third year are also unfavorable, but end up being statistically insignificant once background characteristics are controlled for. Turning next to sickness and unemployment, the effects are small in magnitude and not statistically significant at conventional levels across all years. Offering the intensified treatment bundle instead of standard services appears to have, altogether and on average, no discernible effect on these outcomes. The graphical evidence shown in Figure 2 supports these

²⁰As expected, we find that estimates in Panels A and B are very similar, with standard errors being smaller in the latter.

Table 4: Intention-to-treat effects at different points in time after randomization

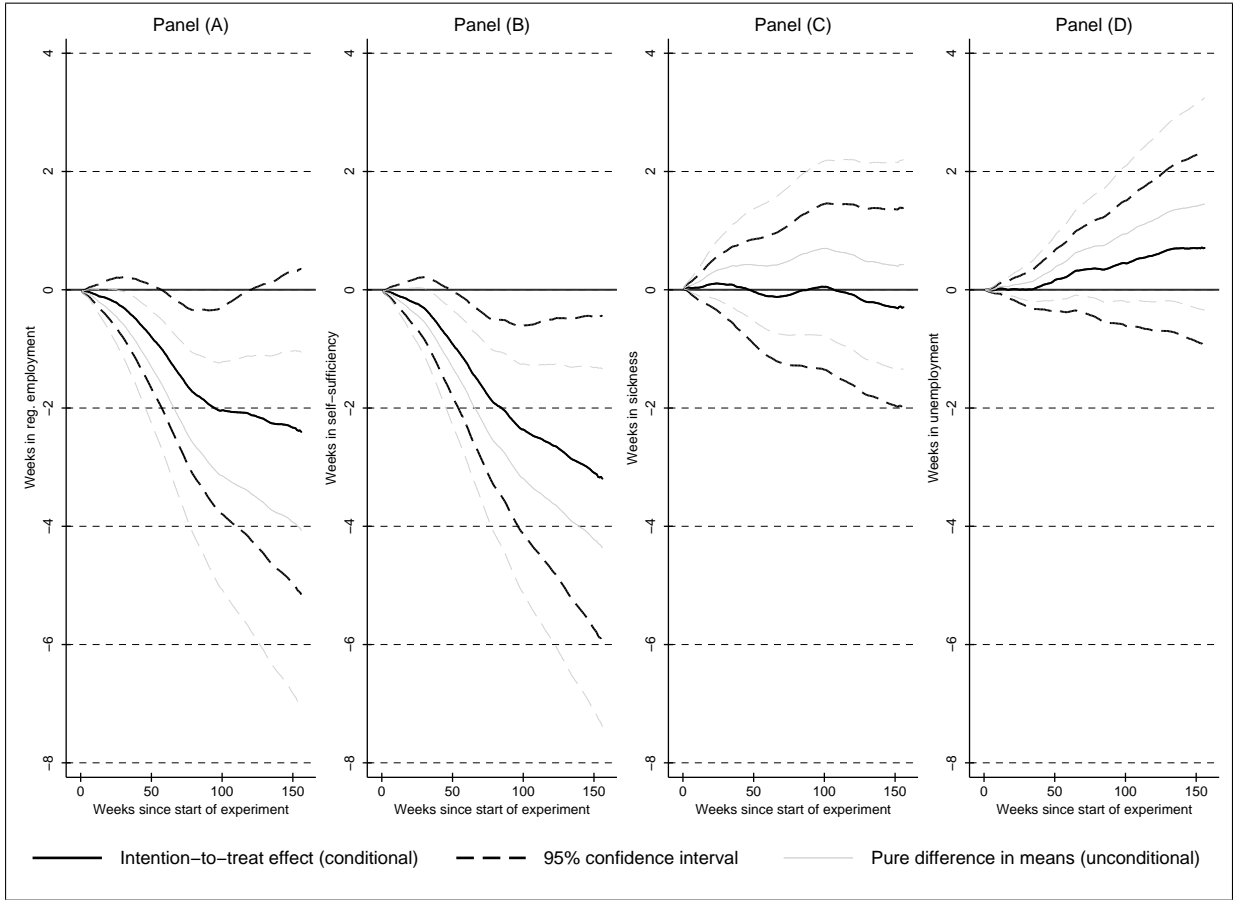
	Dependent variable: Number of weeks in ...			
	I Regular employment	II Self-sufficiency	III Sickness	IV Unemployment
Panel A: Unconditional (not controlling for background characteristics)				
Panel A1: During 1st year after random assignment (N=5,652)				
ITT	-1.390*** (0.505)	-1.395*** (0.516)	0.416 (0.498)	0.409 (0.286)
Mean in control group	16.474*** (0.364)	19.828*** (0.370)	19.812*** (0.356)	4.779*** (0.199)
Panel A2: During 2nd year after random assignment (N=5,652)				
ITT	-1.820*** (0.593)	-1.891*** (0.599)	0.265 (0.394)	0.586 (0.402)
Mean in control group	20.865*** (0.425)	25.774*** (0.426)	8.141*** (0.275)	7.545*** (0.279)
Panel A3: During 3rd year after random assignment (N=5,644)				
ITT	-0.842 (0.607)	-1.065* (0.614)	-0.287 (0.311)	0.436 (0.408)
Mean in control group	20.675*** (0.433)	25.705*** (0.435)	5.197*** (0.223)	7.366*** (0.285)
Panel B: Conditional (controlling for background characteristics)				
Panel B1: During 1st year after random assignment (N=5,652)				
ITT	-0.859* (0.462)	-0.981** (0.471)	-0.048 (0.465)	0.180 (0.269)
Constant	14.138*** (2.639)	19.404*** (3.002)	18.432*** (3.376)	2.168 (1.580)
Panel B2: During 2nd year after random assignment (N=5,652)				
ITT	-1.196** (0.549)	-1.444*** (0.552)	0.075 (0.388)	0.296 (0.377)
Constant	25.333*** (3.069)	37.295*** (3.486)	9.589*** (2.892)	5.056** (2.086)
Panel B3: During 3rd year after random assignment (N=5,644)				
ITT	-0.336 (0.571)	-0.760 (0.572)	-0.341 (0.310)	0.222 (0.385)
Constant	28.774*** (3.200)	44.856*** (3.772)	4.960*** (1.874)	4.874** (2.246)

Notes: The table shows the average causal effect of offering the intensified treatment package as a whole (intention-to-treat). Robust standard errors in parentheses.

Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

findings.

Figure 2: Trajectory of treatment effects (cumulative total intention-to-treat effects)



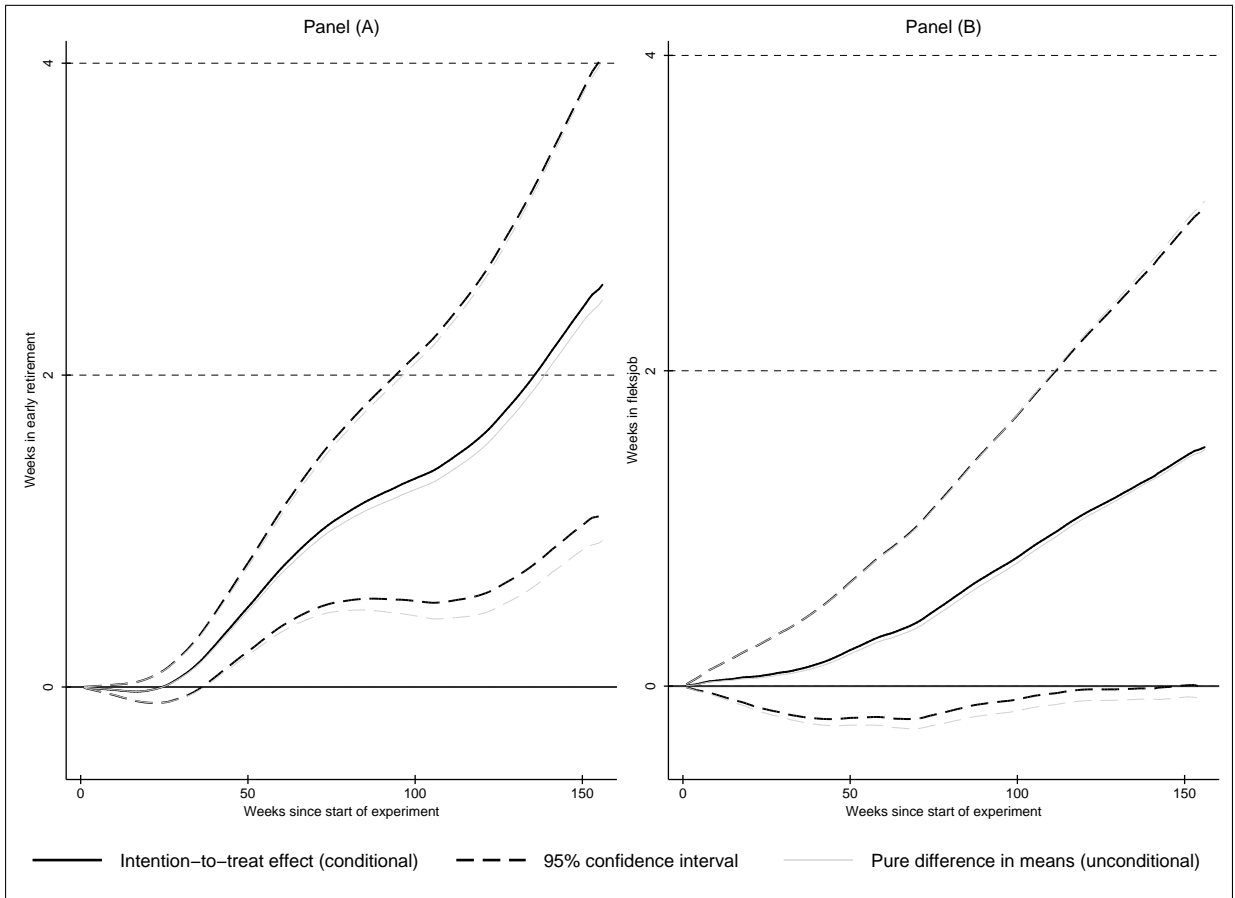
Notes: Each panel is based on 156 separate regressions (one for each week) of the cumulative total number of weeks in (A) regular employment, (B) self-sufficiency, (C) sickness, (D) unemployment on a treatment status dummy and the vector of pre-treatment characteristics summarized in Table 3. The plotted intention-to-treat effects correspond to the difference in the average number of weeks spent in (A) regular employment, (B) self-sufficiency, (C) sickness, (D) unemployment between treated and non-treated individuals (controlling for background characteristics) evaluated at a given point in time after randomization as indicated on the horizontal axes. The dashed lines depict the corresponding pointwise robust confidence intervals at the 95 percent level. The number of observations that each regression is based on varies between 5,652 for all weeks up to and including the 141st week after the start of the experiment (no missings) and 5,644 for the 156th week.

Figure 2 presents a magnified view of the trajectory of treatment effects over time. Each panel is based on 156 separate regressions, one for each week, of the cumulative total number of weeks in the respective labor market status on a treatment status dummy and a vector of background characteristics (comp. above). The time series of treatment effects shown in Panel A illustrates for example that, 156 weeks after random assignment, treated individuals spent on average four weeks less in regular employment than the non-treated (pure difference in means). A causal intention-to-treat effect of about two and a half weeks remains after having partialled out the impact of pre-treatment characteristics; this effect is borderline significant (p-value: 0.087). The adverse impact on non-benefit receipt is even more pronounced and accumulates in a sustained manner over time. The monotonically decreasing cumulative total intention-to-treat effect in Panel B indicates that the self-sufficiency rate among the treated is strictly smaller than for

the non-treated in all weeks. The (negative) gap persists even three years after randomization and contributes to a further decrease in the cumulative total effect. One may conclude that the adverse impact of the treatment goes well beyond an initial locking-in effect. In contrast to that, cumulative total effects on sickness and unemployment are small and not statistically different from zero over the entire domain.

Given that the treatment group's relative *shortfall* in the number of weeks spent in regular employment and other types of self-sufficiency is not fully matched by a relative *abundance* of sickness or unemployment, the question arises into which labor market statuses the treated predominantly transited instead (compared with the non-treated). Figure 3 sheds some light on this by plotting the trajectory of treatment effects for two additional outcomes: early retirement and *fleksjob*. Recall that *fleksjobs* are subsidized jobs targeted at individuals with a permanently reduced work capacity due to a medical condition. *Fleksjob*-workers are typically not expected ever to return to regular working hours. Figure 3 reveals that offering the treatment package promotes transitions into early retirement and *fleksjobs*, unintentionally we presume, for encouraging sick-listed workers to withdraw permanently from productive activities, be it entirely (in the case of early retirement) or in parts (*fleksjobs*), clearly runs against the public interest.

Figure 3: Trajectory of treatment effects for two additional outcome measures (cumulative total intention-to-treat effects)



Notes: See notes to Figure 2.

So far we have confined our discussion to *average* intention-to-treat effects, i.e. to the average effect of offering the treatment among all study participants. However, treatment effects may vary across experimental units for two reasons; first because different subpopulations may respond differently to a given treatment, and second because the composition of actual treatment activities may differ across different subpopulations (see for instance Figure A1 in the Appendix, which shows treatment portfolios by labor market status before sickness). In order to test for effect heterogeneity, we split the sample along two dimensions: we perform subanalyses by labor market status before sickness (regular employment, unemployment, self-employment) and by the degree of benefit dependency in the year prior to assignment (four quartiles). The results, which are shown in Figures A2 and A3 in the Appendix, may be briefly summarized as follows. First, the treatment affects individuals sick-listed from regular employment and individuals sick-listed from unemployment in the same way (adverse impacts on regular employment and self-sufficiency, no effect on sickness and unemployment).²¹ Second, the adverse impact of the treatment is most pronounced for individuals in the first quartile of the benefit dependency distribution in the year before enrolment, i.e., for sick-listed workers with no or only short periods of prior benefit receipt. The effects within this quartile are up to three times greater than for the sample as a whole.

In sum, the treatment package as a whole comes off badly. The treated spent less time in regular employment and other types of self-sufficiency than their peers in the control group. The adverse impact is moderate in magnitude and goes well beyond initial locking-in effects. We find that the treatment is most harmful for individuals with a high degree of self-sufficiency in the year before assignment. Lastly, the treatment promotes transitions into early retirement and *fleksjobs*, both of which are one-way tickets into indefinite periods of welfare dependency, programs of no return, and therefore dead ends on the road to successful reintegration.

5.2 Disentangling the bundle, isolating effects (local treatment strategies approach)

As to the results from the “*local treatment strategy approach*”, have a look at Tables 5 and 6. Table 5 (OLS) displays the average intention-to-treat effects of marginal changes in local treatment strategies. The estimated impacts of actually participating in alternative treatment activities (LATE) are shown in Table 6 (IV/2SLS). Following Markussen and Røed (2014: 19), we have normalized the vectors of local treatment strategy characteristics by scaling its elements φ_{Si} by the inverse of the absolute difference in the average value of φ_{Si} between the local treatment regimes applying the respective treatment activity S least and most. Consequently, a unit difference corresponds to the difference described above and parameter estimates in Table 5

²¹For the subsample of self-employed, we lack statistical power due to the small sample size. Indeed, the effects for this group are very imprecisely estimated.

can be interpreted as the expected change in the outcome variable “resulting from a movement from the treatment environment giving lowest priority to the strategy under consideration to the one giving it highest priority”. Also note that the estimates shown in Table 5 correspond to the *reduced form* estimates of the IV/2SLS model that Table 6 (*second stage*) is based on. The *first stage* is summarized by the F test of excluded instruments reported in the footer of Table 6.

Table 5: Intention-to-treat effects of marginal changes in local treatment strategies at different points in time after randomization (OLS, reduced form estimates)

	Dependent variable: Number of weeks in ...					
	I Regular employ- ment	II Self- sufficiency	III Sickness	IV Unemploy- ment	V Early retirement	VI Fleksjob
Panel A: During 1st year after random assignment ($N=4,728$)						
$\varphi_{\text{traditionalactivation}}$	-2.476* (1.381)	-2.133 (1.401)	-0.444 (1.444)	1.408 (0.859)	0.280 (0.239)	0.768*** (0.268)
$\varphi_{\text{paramedicalcare}}$	-2.937** (1.161)	-2.896** (1.178)	2.073* (1.214)	-0.368 (0.722)	0.119 (0.201)	0.734*** (0.225)
$\varphi_{\text{gradedreturn}}$	4.227*** (1.578)	3.245** (1.602)	-3.538** (1.651)	0.251 (0.981)	0.125 (0.273)	-0.862*** (0.306)
Panel B: During 2nd year after random assignment ($N=4,728$)						
$\varphi_{\text{traditionalactivation}}$	-3.398** (1.656)	-2.610 (1.685)	-0.612 (1.248)	2.077* (1.207)	0.378 (0.655)	1.374** (0.685)
$\varphi_{\text{paramedicalcare}}$	-3.345** (1.392)	-4.055*** (1.417)	1.006 (1.049)	1.304 (1.015)	0.628 (0.550)	1.599*** (0.576)
$\varphi_{\text{gradedreturn}}$	3.674* (1.893)	2.202 (1.927)	-1.446 (1.426)	-2.457* (1.380)	0.622 (0.748)	-1.382* (0.783)
Panel C: During 3rd year after random assignment ($N=4,720$)						
$\varphi_{\text{traditionalactivation}}$	-3.199* (1.707)	-2.065 (1.723)	-0.458 (0.985)	0.515 (1.231)	0.616 (0.935)	1.044 (0.958)
$\varphi_{\text{paramedicalcare}}$	-0.982 (1.436)	-1.552 (1.450)	-0.594 (0.829)	0.652 (1.035)	0.860 (0.787)	1.531* (0.806)
$\varphi_{\text{gradedreturn}}$	3.689* (1.956)	1.906 (1.974)	-0.199 (1.128)	-1.664 (1.410)	0.794 (1.072)	-1.355 (1.098)

Notes: The table shows intention-to-treat effects of marginal changes in local treatment strategies (reduced form estimates). Each panel is based on six separate OLS regressions of the number of weeks in regular (i.e., wage) employment/self-sufficiency/sickness/unemployment/early retirement/fleksjob on the normalized vector of local treatment strategy characteristics and additional controls (individual and municipality-level socio-economic characteristics; week of intake dummies). Standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The OLS estimates in Table 5 suggest that prioritizing graded return-to-work programs has favorable effects. The reduced form estimates indicate that offering these programs more intensively decreases the incidence of sickness during the first year after random assignment. At the same time, they promote transitions into regular employment and other forms of self-sufficiency. The estimates suggest for instance that a movement from the treatment regime prioritizing graded return-to-work programs the least to the one promoting it the most leads to an expected decrease in the time spent in sickness of almost one month (3.5 weeks) during the first year. This decrease is accompanied by a corresponding increase in regular employment and non-benefit receipt. We find that the favorable effect on regular employment persists during the second and

third year after enrolment. During these later years, the impacts on self-sufficiency and sickness are also beneficial, but not statistically significant. While the impact on early retirement cannot be distinguished from zero, prioritizing graded return-to-work programs reduces the likelihood of getting trapped into permanent welfare dependency by taking up subsidized employment under a *fleksjob* arrangement. In sum, graded return-to-work programs are strongly associated with improved labor market prospects. In stark contrast, giving priority to traditional activation and paramedical care impairs subsequent performance. Both of these treatment types exert a negative impact on the incidence of regular employment and non-benefit receipt. In addition, offering these programs more intensively appears to push sick-listed workers into *fleksjobs*. The instrumental variable estimates reported in Table 6 support these findings.

Table 6: Effects of participating in alternative treatment activities (LATE) evaluated at different points in time after randomization (IV/2SLS)

	Dependent variable: Number of weeks in ...					
	I Regular employ- ment	II Self- sufficiency	III Sickness	IV Unemploy- ment	V Early retirement	VI Fleksjob
Panel A: During 1st year after random assignment (N=4,728)						
Traditional activation	-3.333 (2.876)	-2.529 (2.764)	-1.556 (2.985)	2.796* (1.629)	0.403 (0.453)	1.066* (0.627)
Paramedical care	-7.084** (2.761)	-6.551** (2.654)	5.316* (2.866)	-0.794 (1.564)	0.141 (0.434)	1.674*** (0.602)
Graded return	30.290** (11.980)	23.777** (11.515)	-24.260* (12.435)	1.425 (6.787)	0.639 (1.885)	-6.355** (2.610)
Panel B: During 2nd year after random assignment (N=4,728)						
Traditional activation	-4.620 (3.252)	-2.359 (3.174)	-1.539 (2.386)	3.396 (2.410)	0.116 (1.262)	1.672 (1.427)
Paramedical care	-7.510** (3.122)	-8.063*** (3.048)	2.458 (2.291)	3.399 (2.313)	0.790 (1.212)	3.416** (1.370)
Graded return	27.285** (13.546)	17.971 (13.224)	-9.886 (9.941)	-17.532* (10.038)	3.474 (5.258)	-10.530* (5.946)
Panel C: During 3rd year after random assignment (N=4,720)						
Traditional activation	-5.960* (3.374)	-3.088 (3.256)	-0.377 (1.859)	0.830 (2.361)	0.350 (1.801)	1.110 (1.898)
Paramedical care	-3.466 (3.255)	-3.606 (3.140)	-0.931 (1.793)	1.942 (2.277)	1.101 (1.737)	3.314* (1.831)
Graded return	25.984* (14.172)	14.294 (13.674)	-0.722 (7.808)	-11.517 (9.916)	4.372 (7.564)	-10.294 (7.973)

Notes: The table shows the estimated effects of participating in alternative treatment activities (local average treatment effects). Each panel is based on six separate IV/2SLS regressions (second stage). See notes to Table 5 for a description of the outcome variables. Included instruments: see Table A2 and Table A3 in the Appendix; week of intake dummies. Excluded instruments: vector of local treatment strategy characteristics (φ_i). F test of excluded instruments (first stage): traditional activation: 228.75; paramedical care: 480.75; graded return: 20.28. Standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The table shows the estimated effects of participating in alternative programs. It therefore comes as no surprise that these statistics (LATE) tend to be greater in magnitude than those reported in Table 5 (ITT). The estimates in Table 6 may be summarized as follows. First, both

traditional activation and paramedical care seem to have adverse impacts on subsequent labor market outcomes. Participation in a traditional activation program promotes unemployment during the first year after enrolment and is clearly not helping sick-listed workers to reintegrate into the regular labor market. Similarly, traditional activation programs are ineffective in reducing the degree of welfare dependency. The effects of being exposed to paramedical care are even more detrimental. Our results indicate that participating in such a program reduces the time spent in regular employment and self-sufficiency by about one and a half to two months during both the first and second year after intake. These effects are highly statistically significant. And second, letting sick-listed workers participate in graded return-to-work programs is shown to be a very successful strategy. Participation in these programs increases the time spent in regular employment and non-benefit receipt substantially. At the same time, we find favorable effects on sickness during the first year and on unemployment during the second year.

The 4,728 experimental units to which Table 6 refers share an important commonality: they were sick in the (individual) week of intake. However, we must not ignore the fact that the group is diverse in other ways. It is composed of individuals from fundamentally different backgrounds and experiences. It stands to reason that different backgrounds imply different needs, which in turn demand different means: one size *may not* fit all. We made earlier reference to the idea that the sample can, for the purpose of separate analysis, be divided into three groups: individuals sick-listed from regular employment, unemployment, and self-employment. Estimates of local average treatment effects by labor market attachment before sick leave are reported in Table A4.²² The general impression is that the “full sample results” shown in Table 6 are entirely driven by the large subsample of workers sick-listed from regular employment. Treatment effects for the subsample of unemployed individuals cannot be distinguished from zero (only one estimate is statistically significant at conventional levels). The finding that graded return-to-work is ineffective for sick-listed unemployed workers comes as no surprise, given that unemployed workers typically have no work to return to (see also Figure A1 in the Appendix).

A second important dimension along which the study participants vary is their medical condition. The range of conditions spans from chronic pain, cardiovascular diseases, and cancer, to mental health problems like stress, depression, and anxiety (see Table A2 for a complete list). Beyond doubt, different conditions come with varying degrees of lost work capacity, which in turn may call for different cures. At a somewhat crude level of aggregation one can split the sample into two groups: individuals suffering from a mental disorder and workers with a non-mental illness. Table A5 shows local average treatment effects for these groups. For sick-listed workers with a non-mental condition, graded return-to-work programs perform best. While the effects

²²Results for the small subsample of self-employed are suppressed because estimates are too imprecise for the findings to be considered conclusive.

of these programs are generally positive, although imprecisely measured in some specifications, paramedical care and traditional activation appear to have either no or even adverse impacts. These findings correspond very well with the “full sample results” displayed in Table 6. Taken together the empirical evidence strongly advocates the use of graded return-to-work programs for workers sick-listed from regular employment as a result of a non-mental medical condition. The picture for individuals with a psychiatric disorder is less clear. In fact, since none of the estimated effects exceeds its standard error, the results for this group are entirely inconclusive.

5.3 Robustness and sensitivity

Recall that the main idea of the local treatment strategy approach lies in exploiting the variation in treatment priorities across distinct treatment environments in order to construct estimates of the effects of alternative rehabilitation programs on subsequent labor market outcomes. In a first step, treatment strategies are proxied by vectors of local treatment strategy characteristics, which we estimate in the framework of a competing risks transition rate model. In a second step, these proxies are used as instruments for actual treatment events. We earlier argued that local treatment strategy characteristics provide legitimate instruments for the potentially endogenous program participation indicators (see Section 4.2). A key assumption is the *independence* of instruments and potential outcomes: for the individual sick-listed worker, being exposed to a particular local treatment regime is *as good as randomly assigned*, and so are the instruments. As a matter of fact, much of the variation in treatment strategies is *completely random* because of the experimental intervention (the variation between treatment and control group *within* a given job center is truly random). This is where the present study differs from earlier research that relies on this identification strategy, above all the observational study by Markussen and Røed (2014).

Given this setup, we consider three robustness tests. First, we isolate the truly random part of the variation in treatment strategies by redefining the notion of a “local treatment environment”. Up to this point, a local treatment environment has been defined to consist of a particular job center-treatment status cell: the control group in Copenhagen constitutes for instance *one* local treatment environment. This definition allows us to exploit the variation in treatment intensities both *between* the 16 job centers covered by the experiment and between treatment and control group *within* a given job center (32 local treatment environments). Yet, this approach has a slight drawback: only part of the variation in treatment strategies is *completely* random (as discussed above). Fortunately, the truly random part can easily be isolated by letting a local treatment environment correspond to an individual’s treatment status (two truly randomly assigned local treatment environments). The advantage of this approach is that it rules out purposeful sorting into treatment regimes and thus guarantees *independence*. Table 7 shows the results of this

robustness test. It is immediately apparent that our main findings (Table 6) are robust to this alternative approach. The results are the same as before, both in terms of the direction of the effects and their statistical significance (in fact, the estimates presented here are even more precise than those in Table 6). In particular, we again find that graded return-to-work programs are associated with favorable effects, while traditional activation and paramedical care perform poorly.²³

Table 7: Effects of participating in alternative treatment activities (LATE) evaluated at different points in time after randomization accounting only for truly random variation between treatment and control group (IV/2SLS)

	Dependent variable: Number of weeks in ...					
	I Regular employ- ment	II Self- sufficiency	III Sickness	IV Unemploy- ment	V Early retirement	VI Fleksjob
Panel A: During 1st year after random assignment ($N=4,728$)						
Traditional activation	-2.720** (1.114)	-2.706** (1.133)	-0.644 (1.195)	1.413** (0.703)	0.065 (0.197)	0.790*** (0.221)
Paramedical care	-3.529*** (1.262)	-4.032*** (1.283)	1.229 (1.353)	0.592 (0.796)	0.366 (0.223)	0.418* (0.250)
Graded return	10.606*** (0.899)	11.005*** (0.915)	-8.124*** (0.965)	-2.890*** (0.567)	-0.124 (0.159)	0.083 (0.178)
Panel B: During 2nd year after random assignment ($N=4,728$)						
Traditional activation	-2.126 (1.338)	-2.857** (1.364)	-1.332 (1.025)	0.119 (0.988)	0.211 (0.537)	2.180*** (0.561)
Paramedical care	-5.118*** (1.516)	-6.193*** (1.544)	1.811 (1.161)	1.077 (1.119)	1.242** (0.608)	0.871 (0.635)
Graded return	9.216*** (1.080)	8.701*** (1.101)	-3.355*** (0.827)	-3.496*** (0.798)	-1.576*** (0.434)	0.210 (0.453)
Panel C: During 3rd year after random assignment ($N=4,720$)						
Traditional activation	-1.178 (1.391)	-2.349* (1.406)	-0.502 (0.811)	-1.215 (1.010)	0.112 (0.768)	2.564*** (0.786)
Paramedical care	-2.833* (1.572)	-3.633** (1.590)	-0.613 (0.917)	1.251 (1.142)	2.445*** (0.868)	0.880 (0.889)
Graded return	8.091*** (1.122)	7.025*** (1.134)	-0.517 (0.654)	-3.519*** (0.814)	-2.368*** (0.619)	-0.146 (0.634)

Notes: First robustness test. The table shows the estimated effects of participating in alternative treatment activities (local average treatment effects). Each panel is based on six separate IV/2SLS regressions (second stage). See notes to Table 5 for a description of the outcome variables. Included instruments: see Table A2 and Table A3 in the Appendix; week of intake dummies. Excluded instruments: vector of local treatment strategy characteristics (φ_i). F test of excluded instruments (first stage): traditional activation: 2,007.29; paramedical care: 839.18; graded return: 2,837.23. Standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

As a second test of robustness, let us pretend that the experiment did not take place. In the absence of the experimental intervention, all sick-listed workers would have received treatment as usual, i.e., “control group treatment”, and for this second robustness test, we only use “control group data”. One can consider this robustness check to be the analysis we would have undertaken

²³Note that the statistics shown in Table 7 are of smaller magnitude than those reported in Table 6. It is worth noting that this does not cast doubt on the reliability of the estimates. By changing the way a local treatment environment is defined, we end up with a different set of instruments, which may be associated with a different compliant subpopulation and thus different LATEs (Angrist and Pischke, 2008).

Table 8: Effects of participating in alternative treatment activities (LATE) evaluated at different points in time after randomization using control group data only (IV/2SLS)

	Dependent variable: Number of weeks in ...					
	I Regular employ- ment	II Self- sufficiency	III Sickness	IV Unemploy- ment	V Early retirement	VI Fleksjob
Panel A: During 1st year after random assignment (N=2,304)						
Traditional activation	1.266 (24.452)	-21.257 (24.599)	-11.582 (26.406)	19.868 (15.529)	0.110 (3.636)	3.680 (3.885)
Paramedical care	-42.899 (28.897)	-50.258* (29.070)	54.086* (31.206)	6.216 (18.352)	-2.025 (4.296)	-1.363 (4.591)
Graded return	18.991 (12.080)	8.493 (12.152)	-16.335 (13.044)	8.698 (7.671)	0.134 (1.796)	-2.754 (1.919)
Panel B: During 2nd year after random assignment (N=2,304)						
Traditional activation	-34.025 (28.503)	-36.207 (29.665)	-13.958 (22.464)	27.092 (22.195)	5.141 (11.047)	10.010 (10.604)
Paramedical care	-12.943 (33.684)	-39.734 (35.057)	35.473 (26.547)	-21.887 (26.229)	20.719 (13.055)	2.284 (12.532)
Graded return	6.178 (14.080)	-0.098 (14.655)	-15.939 (11.097)	13.628 (10.964)	0.860 (5.457)	-4.076 (5.239)
Panel C: During 3rd year after random assignment (N=2,297)						
Traditional activation	-19.581 (29.738)	-6.528 (29.972)	-19.304 (19.554)	-0.950 (21.476)	10.070 (17.719)	23.426 (16.647)
Paramedical care	-5.076 (34.357)	-24.484 (34.628)	6.468 (22.592)	-4.116 (24.811)	38.658* (20.472)	-5.539 (19.233)
Graded return	11.914 (14.783)	12.720 (14.899)	-15.767 (9.721)	4.051 (10.676)	7.506 (8.808)	-3.900 (8.275)

Notes: Second robustness test. The table shows the estimated effects of participating in alternative treatment activities (local average treatment effects). Each panel is based on six separate IV/2SLS regressions (second stage). See notes to Table 5 for a description of the outcome variables. Included instruments: see Table A2 and Table A3 in the Appendix; week of intake dummies. Excluded instruments: vector of local treatment strategy characteristics (φ_i). F test of excluded instruments (first stage): traditional activation: 7.14; paramedical care: 8.34; graded return: 9.24. Standard errors in parentheses.

Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

were this an observational study like that of Markussen and Røed (2014). At the same time, assuming away the trial is a neat way of showing that we are—in our original analysis—benefiting from the additional truly random variation in treatment strategies induced by the experiment. Now, recall that control group members are distributed across 16 job centers (Table A1 in Appendix). Hence we now focus on the variation in treatment strategies across 16 local treatment environments. The results of this are shown in Table 8. Note first that the effects reported here tend to be in the same direction as our main results: only every fifth coefficient changed sign compared to the benchmark (Table 6). However, few of the estimates displayed in Table 8 are statistically significant at conventional levels. For one thing, the precision loss might be caused by the sample size reduction (given that we are now using only “control group data”). For another, it may arise from the reduced extent of exploitable variation in local treatment strategies. Given these considerations, one might interpret the results shown in Table 8 in the following way: while the direction of the effects is (by and large) accurately measured, the variation in treatment strategies in the control group is too small to derive statistically significant estimates, given the relatively small sample size. Precision could be increased either by increasing the sample size (a large scale observational study, like Markussen and Røed (2014)) or by exploiting additional truly random variation in treatment strategies (as we do in our original analysis). Viewed this way, the results from this second robustness test serve a dual purpose. First, they substantiate our previous findings, and second, they illustrate that we are benefiting from the additional truly random variation in treatment strategies—a unique characteristic of the paper at hand.

The third robustness check is intended to test whether our results still apply when an alternative set of instruments is used. The instruments used so far were derived from a competing risks transition rate model: we used the (leave-out) mean covariate-adjusted transition propensities (into alternative treatment activities) of all sick-listed workers within a particular local treatment environment as instruments for program participation (see Section 4.2). The instruments used in this robustness test are simpler to construct. To construct the instruments, we estimate—separately for each of the three alternative treatments—the following linear probability model (one observation per sick-listed worker):

$$P_{Si} = \beta_0 + \mathbf{X}_i\boldsymbol{\eta}_S + \mathbf{X}_j\boldsymbol{\kappa}_S + \mathbf{C}\boldsymbol{\rho}_S + u_{Sij} \quad (7)$$

where P_{Si} (S =traditional activation, paramedical care, graded return-to-work) are indicators of actual treatment receipt (P_{Si} equals unity in case of participation in treatment activity S , zero otherwise); \mathbf{X}_i is a vector of individual pre-treatment characteristics; \mathbf{X}_j is a vector of municipality-level controls; and \mathbf{C} is a vector of “center dummies”—one for each local treatment environment. Note that the (32) elements of $\boldsymbol{\rho}_S$ can be interpreted as covariate-adjusted indica-

tors of a local treatment environment’s inclination to use treatment S . Therefore, ρ_S provides a natural proxy for intrinsically unobserved local treatment cultures of different treatment environments. Pursuant to the rationale discussed in Section 4.2, it can be argued that these proxies for “local area practice styles” provide legitimate instruments for the potentially endogenous program participation indicators. Table 9 below reports local average treatment effects estimated by using ρ_S as instrument for P_{Si} (see equation (6) for the structural model). It can immediately be seen that the estimates reported here correspond quite well with our main results (Table 6).

Table 9: Effects of participating in alternative treatment activities (LATE) evaluated at different points in time after randomization using center fixed effects as instruments for treatment participation (IV/2SLS)

	Dependent variable: Number of weeks in ...					
	I Regular employ- ment	II Self- sufficiency	III Sickness	IV Unemploy- ment	V Early retirement	VI Fleksjob
Panel A: During 1st year after random assignment (N=4,688)						
Traditional activation	-3.174 (2.334)	-3.280 (2.354)	-1.178 (2.510)	2.628* (1.464)	0.449 (0.409)	0.782* (0.471)
Paramedical care	-3.852** (1.747)	-3.858** (1.762)	2.917 (1.879)	-0.058 (1.095)	0.158 (0.306)	0.758** (0.353)
Graded return	14.388** (5.835)	12.009** (5.885)	-12.196* (6.275)	-0.559 (3.659)	0.225 (1.022)	-2.177* (1.178)
Panel B: During 2nd year after random assignment (N=4,688)						
Traditional activation	-3.553 (2.829)	-1.708 (2.839)	-2.171 (2.128)	1.780 (2.126)	0.626 (1.116)	1.196 (1.168)
Paramedical care	-5.799*** (2.117)	-6.485*** (2.125)	1.089 (1.592)	3.785** (1.591)	1.049 (0.835)	1.405 (0.875)
Graded return	18.750*** (7.072)	12.195* (7.097)	-1.520 (5.319)	-14.042*** (5.314)	-0.040 (2.790)	-3.251 (2.922)
Panel C: During 3rd year after random assignment (N=4,680)						
Traditional activation	-5.012* (2.932)	-2.692 (2.932)	-0.934 (1.702)	-0.201 (2.116)	0.894 (1.598)	0.579 (1.639)
Paramedical care	-1.216 (2.178)	-1.522 (2.177)	-1.432 (1.264)	1.897 (1.571)	1.976* (1.187)	1.160 (1.217)
Graded return	15.302** (7.266)	7.587 (7.264)	3.395 (4.217)	-7.642 (5.244)	-2.345 (3.960)	-2.314 (4.061)

Notes: Third robustness test. The table shows the estimated effects of participating in alternative treatment activities (local average treatment effects). Each panel is based on six separate IV/2SLS regressions (second stage). See notes to Table 5 for a description of the outcome variables. Included instruments: see Table A2 and Table A3 in the Appendix; week of intake dummies. Excluded instruments: center fixed effects; one center had to be excluded due to perfect multicollinearity in the linear probability model used to construct the instruments. F test of excluded instruments (first stage): traditional activation: 251.40; paramedical care: 540.22; graded return: 33.26. Standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

6 Conclusion

This paper has provided new and important evidence relating to the effects of intensive mandatory return-to-work activities on sick-listed workers’ subsequent labor market outcomes. We have used a unique dataset from a large-scaled randomized experiment conducted in Danish job

centers in 2009, linked to large administrative registers.

We first evaluated the *intention-to-treat* effect of the intensified activation package as a whole (weekly meetings with caseworkers combined with graded return-to-work, traditional activation and/or paramedical care) by simply comparing the average labor market outcomes of treated and non-treated (“difference in means”). Second, we have exploited variations in local treatment strategies, both between job centers and between treatment and control groups within a given job center—in spirit of Markussen and Røed (2014)—to compare the relative effectiveness of the alternative measures.

Our findings reveal first that the experimental intervention as a whole has been ineffective. Sick-listed workers initially assigned to the treatment group spent less time in regular employment and self-sufficiency (i.e., all forms of non-benefit receipt) compared to their peers in the control group who benefited from the usual forms of intervention. Nevertheless, our results also show that offering graded return-to-work programs more intensively is associated with an increase in regular employment and self-sufficiency, and a decrease in unemployment. Traditional activation and paramedical care, on the other hand, appear to have either no or even adverse impacts.

Taken together, our results thus suggest that graded return-to-work programs are the most effective intervention for improving sick-listed workers’ subsequent labor outcomes. When intensified, they are associated with strong and lasting effects, for sick-listed workers who were employed before becoming sick and who do not suffer from mental disorders. Therefore, our results support a greater reliance on graded return-to-work programs to help sick-listed individuals to return to work as quickly as possible. Recent medical research also finds in favor of an increased emphasis on graded return-to-work (see e.g., Waddell and Burton (2006)).

References

- Aakvik A., Heckman J., and Vytlacil E. (2005). Estimating Treatment Effects for Discrete Outcomes when Responses to Treatment Vary: An Application to Norwegian Vocational Rehabilitation Programs. *Journal of Econometrics*, 125(1-2): 15–51.
- Andrén D. and Svensson M. (2012). Part-Time Sick Leave as a Treatment Method for Individuals with Musculoskeletal Disorders. *Journal of Occupational Rehabilitation*, 22(3): 418–426.
- Angrist J. and Pischke J.S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.
- Boll J., Hertz M., Rosholm M., and Svarer M. (2010). Evaluering af Aktive - Hurtigere Tilbage. Evaluering, Arbejdsmarkedsstyrelsen.
- Card D., Kluve J., and Weber A. (2010). Active Labour Market Policy Evaluations: A Meta-Analysis. *Economic Journal*, 120(548): 452–477.
- Engström P., Hägglund P., and Johansson P. (2015). Early Interventions and Disability Insurance: Experience from a Field Experiment. *Forthcoming in The Economic Journal*.
- Everhardt T. and de Jong P. (2011). Return to Work After Long Term Sickness. *De Economist*, 159(3): 361–380.
- Frölich M., Heshmati A., and Lechner M. (2004). A Microeconomic Evaluation of Rehabilitation of Long-term Sickness in Sweden. *Journal of Applied Econometrics*, 19(3): 375–396.
- Hesselius P. (2007). Does Sickness Absence Increase the Risk of Unemployment? *The Journal of Socio-Economics*, 36(2): 288–310.
- Høgelund J., Holm A., and Eplov L. (2012). The Effect of Part-time Sick Leave for Employees with Mental Disorders. *Journal of Mental Health Policy and Economics*, 15(4): 157–170.
- Høgelund J., Holm A., and McIntosh J. (2010). Does Graded Return-to-work Improve Sick-listed Workers' Chance of Returning to Regular Working Hours? *Journal of Health Economics*, 29(1): 158–169.
- Imbens G. and Angrist J. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2): 467–75.
- Kausto J., Miranda H., Martimo P., and Viikari-Juntura E. (2008). Partial Sick Leave: Review of Its Use, Effects and Feasibility in the Nordic Countries. *Scandinavian Journal of Work, Environment and Health*, 34(4): 239–249.

- Markussen S. (2012). The Individual Cost of Sick Leave. *Journal of Population Economics*, 25(4): 1287–1306.
- Markussen S., Mykletun A., and Røed K. (2012). The Case for Presenteeism - Evidence from Norway's Sickness Insurance Program. *Journal of Public Economics*, 96(11): 959–972.
- Markussen S. and Røed K. (2014). The Impacts of Vocational Rehabilitation. *Labour Economics*, 31: 1–13.
- Nielsen M.B., Hansen J.V., Aust B., Tverborgvik T., Thomsen B., Bjorner J.B., Mortensen O.S., Rugulies R., Winzor G., Ørbæk P., Helverskov T., Kristensen N., and Poulsen O.M. (2014). A Multisite Randomized Controlled Trial on Time to Self-support among Sickness Absence Beneficiaries. The Danish National Return-to-work Programme. *European Journal of Public Health*.
- OECD (2010). *Sickness, Disability and Work: Breaking the Barriers*.
- Van Oostrom S., Driessen M., de Vet H., Franche R., Schonstein E., Loisel P., Mechelen W., and Anema J. (2009). *Workplace Interventions for Preventing Work Disability (Review)*. The Cochrane Library. John Wiley & Sons, Ltd.
- Viikari-Juntura E., Kausto J., Shiri R., Kaila-Kangas L., Takala E., Karppinen J., Miranda H., Luukkonen R., and Martimo P. (2012). Return to Work After Early Part-time Sick Leave due to Musculoskeletal Disorders: a Randomized Controlled Trial. *Scandinavian Journal of Work, Environment and Health*, 38(2): 134–143.
- Waddell G. and Burton A. (2006). *Is Work Good for Your Health and Well-being?* TSO (The Stationery Office), London.

Appendix A

Table A1: Number of individuals by employment region, job center and treatment status

Employment region, job center	Control group	Treatment group	Total
<i>Panel A: Region Zealand (N=1,902)</i>			
Bornholm	177	212	389
Gentofte	48	49	97
Greve	204	237	441
København	206	250	456
Ringsted	174	191	365
Vordingborg	73	81	154
<i>Panel B: North Denmark Region (N=525)</i>			
Aalborg	210	215	425
Morsø	50	50	100
<i>Panel C: Central Denmark Region (N=913)</i>			
Randers	136	114	250
Holstebro	231	116	347
Herning	98	100	198
Horsens	58	60	118
<i>Panel D: Region of Southern Denmark (N=2,312)</i>			
Svendborg	114	167	281
Nyborg	191	185	376
Odense	482	506	988
Aabenraa	343	324	667
Total	2,795	2,857	5,652

Table A2: Pre-treatment characteristics by treatment status (N=4,728)

Variable	Control group		Treatment group		Difference
	Mean	Std. dev.	Mean	Std. dev.	
<i>Panel A: Socio-demographic characteristics</i>					
< 30 years old	0.150	(0.357)	0.137	(0.343)	-0.014
30-39 years old	0.234	(0.423)	0.218	(0.413)	-0.015
40-49 years old	0.273	(0.446)	0.286	(0.452)	0.013
> 49 years old	0.343	(0.475)	0.359	(0.480)	0.016
Male	0.433	(0.496)	0.457	(0.498)	0.024
Married	0.583	(0.493)	0.570	(0.495)	-0.013
Danish origin	0.878	(0.327)	0.876	(0.329)	-0.002
Western origin	0.045	(0.207)	0.050	(0.217)	0.005
Non-western origin	0.077	(0.267)	0.074	(0.262)	-0.003
<i>Panel B: Labor market history</i>					
Sick-listed from regular empl.	0.777	(0.416)	0.743	(0.437)	-0.034***
Sick-listed from UI-benefits	0.160	(0.366)	0.194	(0.395)	0.034***
Sick-listed from self-empl.	0.063	(0.243)	0.063	(0.242)	-0.000
Elapsed sickness duration ^a	8.902	(6.235)	8.601	(6.082)	-0.300*
Time spent on sickness-ben. ^b	0.212	(0.143)	0.208	(0.145)	-0.003
Time spent on sickness-ben. ^c	0.080	(0.176)	0.080	(0.175)	-0.000
Time spent on sickness-ben. ^d	0.069	(0.163)	0.063	(0.156)	-0.006
Degree of pub. inc. support ^{b e}	0.311	(0.251)	0.317	(0.259)	0.006
Degree of pub. inc. support ^{c e}	0.216	(0.314)	0.213	(0.314)	-0.003
Degree of pub. inc. support ^{d e}	0.240	(0.338)	0.232	(0.332)	-0.007
<i>Panel C: Diagnoses</i>					
Back and neck disorders	0.152	(0.359)	0.164	(0.370)	0.011
Moving incapacity	0.196	(0.397)	0.206	(0.405)	0.011
Musculoskeletal disorders	0.038	(0.192)	0.047	(0.213)	0.009
Cardiovascular diseases	0.047	(0.211)	0.048	(0.214)	0.001
Stomach/liver/kidney diseases	0.033	(0.179)	0.033	(0.179)	0.000
Mental health problems	0.334	(0.472)	0.319	(0.466)	-0.015
Chronic pain	0.007	(0.083)	0.009	(0.097)	0.003
Cancer	0.033	(0.177)	0.040	(0.197)	0.008
Other	0.160	(0.367)	0.132	(0.339)	-0.028***
Number of observations	2,304		2,424		

^a at start of experiment (in weeks) ^b in year before sickness ^c in second last year before sickness ^d in third last year before sickness ^e any kind of public income support.

Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A3: Municipality-level controls used both in competing risks model and in main outcome equation (local treatment strategies approach)

Variable	Description
<i>Panel A: Socio-demographic characteristics</i>	
avage	Average age
fertil	Total fertility rate
lifeexp	Life expectancy for new born babies
danes	Persons of Danish origin in % of the population
lowedu	Percentage of working-age population with no more than ten years of schooling
migra1	External net migration in % of the population
migra2	Internal net migration in % of the population
crime	Reported criminal offences per capita
<i>Panel B: Local labor market conditions</i>	
unemp	Full-time unemployed persons in % of the labor force
sbens1	Persons receiving sickness benefits (w/ job) in % of the working-age population
sbens2	Persons receiving sickness benefits (w/o job) in % of the working-age population
lfpact	Economic activity rate (labor force participation)
outlf	Percentage of working-age population outside the labor force
prisec	Persons employed in primary sector in % of the population
vacs1	Number of open vacancies per unemployed
vacs2	Number of newly opened vacancies per unemployed

Notes: Data on all variables (except for *vacs1* and *vacs2*) stem from Statistics Denmark, Denmark's national statistics institute.

Table A4: Effects of participating in alternative treatment activities (LATE) evaluated at different points in time after randomization by labor market status before sickness (IV/2SLS)

	Dependent variable: Number of weeks in ...					
	I Regular employ- ment	II Self- sufficiency	III Sickness	IV Unemploy- ment	V Early retirement	VI Fleksjob
Panel A: Labor market status before sickness: Regular employment						
Panel A1: During 1st year after random assignment (N=3,593)						
Traditional activation	-4.896 (3.932)	-3.413 (3.741)	-1.712 (3.964)	3.438* (1.852)	0.552 (0.526)	1.375 (0.883)
Paramedical care	-9.990** (4.138)	-9.896** (3.936)	8.836** (4.171)	-2.060 (1.948)	0.318 (0.554)	2.449*** (0.930)
Graded return	36.255** (14.543)	31.223** (13.833)	-32.384** (14.660)	3.446 (6.848)	0.650 (1.945)	-7.333** (3.267)
Panel A2: During 2nd year after random assignment (N=3,593)						
Traditional activation	-4.193 (4.103)	-0.287 (3.987)	-2.460 (2.901)	2.568 (2.610)	-0.242 (1.535)	2.166 (1.833)
Paramedical care	-9.454** (4.318)	-10.832*** (4.196)	4.597 (3.053)	1.749 (2.747)	0.706 (1.615)	4.447** (1.929)
Graded return	26.203* (15.174)	20.595 (14.746)	-13.074 (10.728)	-9.846 (9.654)	4.192 (5.675)	-12.261* (6.780)
Panel A3: During 3rd year after random assignment (N=3,588)						
Traditional activation	-7.247* (4.352)	-3.890 (4.318)	-0.365 (2.181)	1.337 (2.667)	-0.603 (2.160)	1.912 (2.499)
Paramedical care	-4.892 (4.605)	-7.580* (4.569)	0.389 (2.308)	-1.388 (2.822)	1.236 (2.286)	5.278** (2.645)
Graded return	28.269* (16.264)	26.934* (16.134)	-4.043 (8.152)	-1.991 (9.967)	4.474 (8.071)	-16.558* (9.340)
Panel B: Labor market status before sickness: Unemployment						
Panel B1: During 1st year after random assignment (N=838)						
Traditional activation	-0.681 (3.288)	-0.329 (3.535)	1.857 (5.200)	1.526 (4.220)	-0.580 (1.210)	0.087 (0.590)
Paramedical care	-1.579 (3.286)	-1.604 (3.533)	-0.517 (5.196)	-2.321 (4.217)	0.358 (1.210)	0.762 (0.589)
Graded return	0.464 (48.671)	-16.020 (52.337)	44.974 (76.974)	26.351 (62.464)	-18.111 (17.917)	-8.015 (8.730)
Panel B2: During 2nd year after random assignment (N=838)						
Traditional activation	-6.925 (4.761)	-4.427 (5.515)	4.205 (4.841)	2.024 (5.601)	-0.856 (2.566)	-0.124 (2.625)
Paramedical care	2.498 (4.758)	0.874 (5.512)	-3.581 (4.838)	0.057 (5.597)	0.196 (2.564)	4.346* (2.623)
Graded return	-44.981 (70.473)	-71.754 (81.641)	19.302 (71.669)	46.755 (82.910)	5.994 (37.981)	-22.836 (38.854)
Panel B3: During 3rd year after random assignment (N=835)						
Traditional activation	-5.403 (5.213)	-2.365 (5.454)	4.992 (5.258)	-6.319 (5.641)	2.117 (3.976)	1.954 (3.468)
Paramedical care	4.302 (5.189)	-0.143 (5.429)	-7.702 (5.234)	9.036 (5.616)	-1.722 (3.958)	2.705 (3.452)
Graded return	-59.668 (76.900)	-58.421 (80.452)	90.707 (77.554)	-63.715 (83.215)	42.490 (58.647)	-1.651 (51.155)

Notes: Standard errors in parentheses.

Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

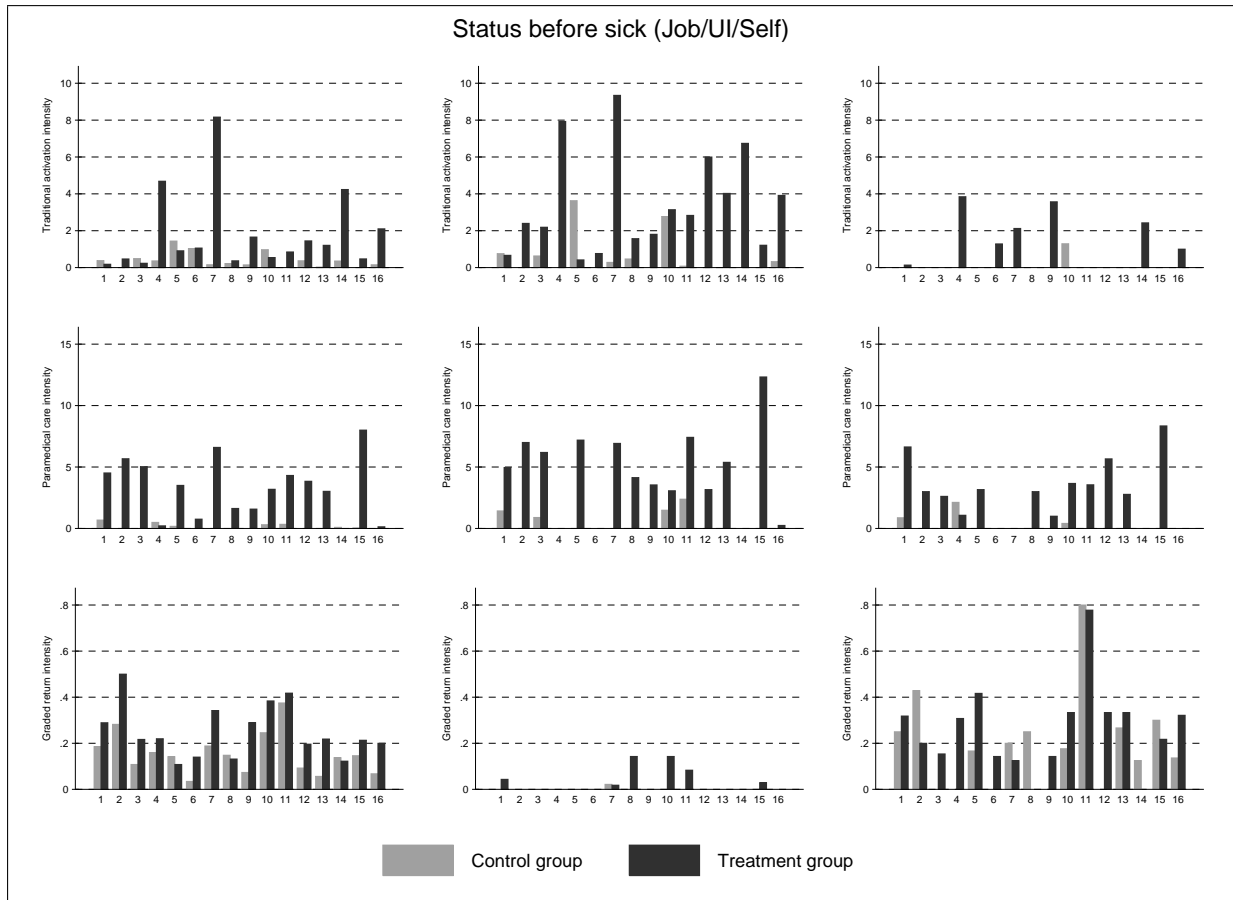
Table A5: Effects of participating in alternative treatment activities (LATE) evaluated at different points in time after randomization by sickness type (IV/2SLS)

	Dependent variable: Number of weeks in ...					
	I Regular employ- ment	II Self- sufficiency	III Sickness	IV Unemploy- ment	V Early retirement	VI Fleksjob
Panel A: Sickness type: Non-mental sickness						
Panel A1: During 1st year after random assignment (N=3,184)						
Traditional activation	-4.275 (3.802)	-4.743 (3.776)	1.588 (3.881)	2.321 (1.971)	0.791 (0.617)	1.658 (1.036)
Paramedical care	-6.560* (3.501)	-6.144* (3.477)	4.777 (3.574)	-1.789 (1.815)	0.005 (0.568)	2.129** (0.954)
Graded return	30.801** (15.312)	28.375* (15.208)	-25.392 (15.631)	0.511 (7.936)	1.267 (2.483)	-10.177** (4.170)
Panel A2: During 2nd year after random assignment (N=3,184)						
Traditional activation	-6.049 (4.228)	-6.025 (4.241)	1.812 (3.277)	2.211 (3.071)	0.945 (1.729)	1.949 (1.994)
Paramedical care	-7.342* (3.893)	-7.272* (3.905)	3.759 (3.018)	2.452 (2.828)	0.552 (1.593)	3.205* (1.836)
Graded return	27.197 (17.028)	23.607 (17.081)	-19.251 (13.198)	-19.809 (12.369)	5.410 (6.965)	-12.903 (8.029)
Panel A3: During 3rd year after random assignment (N=3,181)						
Traditional activation	-7.602 (4.647)	-5.802 (4.435)	0.478 (2.573)	1.556 (3.015)	0.897 (2.320)	2.039 (2.470)
Paramedical care	-3.882 (4.273)	-3.129 (4.078)	1.729 (2.366)	1.459 (2.772)	2.473 (2.133)	2.038 (2.271)
Graded return	35.566* (18.597)	26.155 (17.752)	-12.268 (10.297)	-16.986 (12.066)	0.240 (9.283)	-9.422 (9.886)
Panel B: Sickness type: Mental sickness						
Panel B1: During 1st year after random assignment (N=1,544)						
Traditional activation	-1.284 (21.214)	5.830 (21.007)	2.141 (60.340)	-10.680 (72.118)	1.170 (8.751)	2.952 (12.660)
Paramedical care	-8.616 (42.464)	-0.177 (42.048)	28.233 (120.781)	-31.032 (144.360)	4.343 (17.516)	6.200 (25.341)
Graded return	36.272 (219.939)	-16.177 (217.785)	-139.830 (625.578)	165.143 (747.704)	-19.951 (90.724)	-29.125 (131.251)
Panel B2: During 2nd year after random assignment (N=1,544)						
Traditional activation	3.549 (26.561)	18.189 (63.758)	-38.318 (129.939)	21.028 (58.765)	-0.846 (9.985)	11.731 (49.175)
Paramedical care	2.487 (53.166)	18.562 (127.623)	-57.365 (260.097)	27.807 (117.628)	4.719 (19.988)	24.052 (98.432)
Graded return	-25.376 (275.368)	-132.725 (661.013)	303.553 (1347.146)	-137.756 (609.241)	-14.903 (103.524)	-115.386 (509.820)
Panel B3: During 3rd year after random assignment (N=1,539)						
Traditional activation	15.504 (155.994)	24.716 (197.655)	-37.859 (324.129)	-28.692 (270.304)	-22.179 (186.662)	30.706 (289.803)
Paramedical care	32.379 (318.409)	38.549 (403.438)	-76.287 (661.598)	-59.552 (551.730)	-40.507 (381.007)	68.841 (591.543)
Graded return	-172.082 (1,668.178)	-225.795 (2,113.648)	391.907 (3,466.182)	319.714 (2,890.555)	221.979 (1,996.136)	-348.506 (3,099.160)

Notes: Standard errors in parentheses.

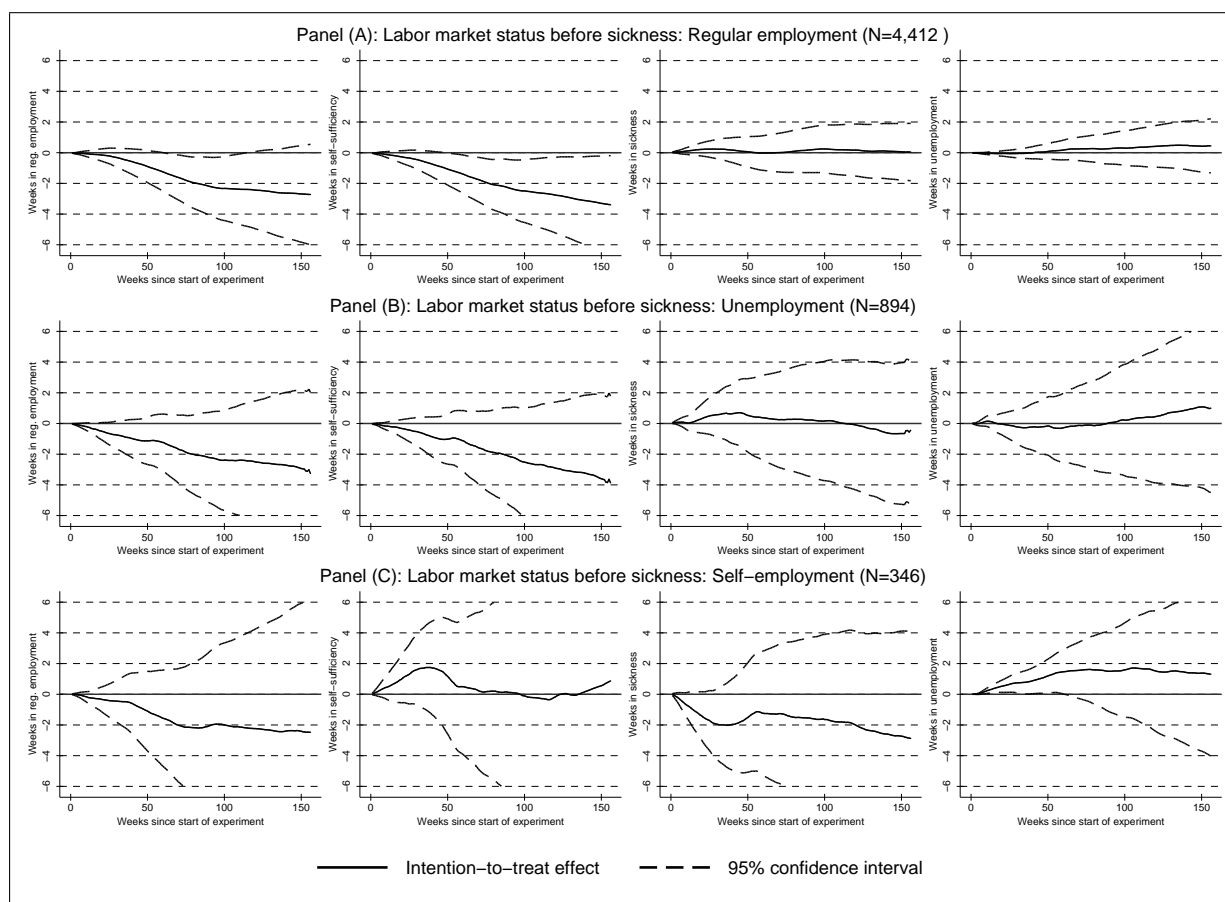
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure A1: Traditional activation, paramedical care and graded return-to-work intensities by job center, treatment status and labor market status before sickness



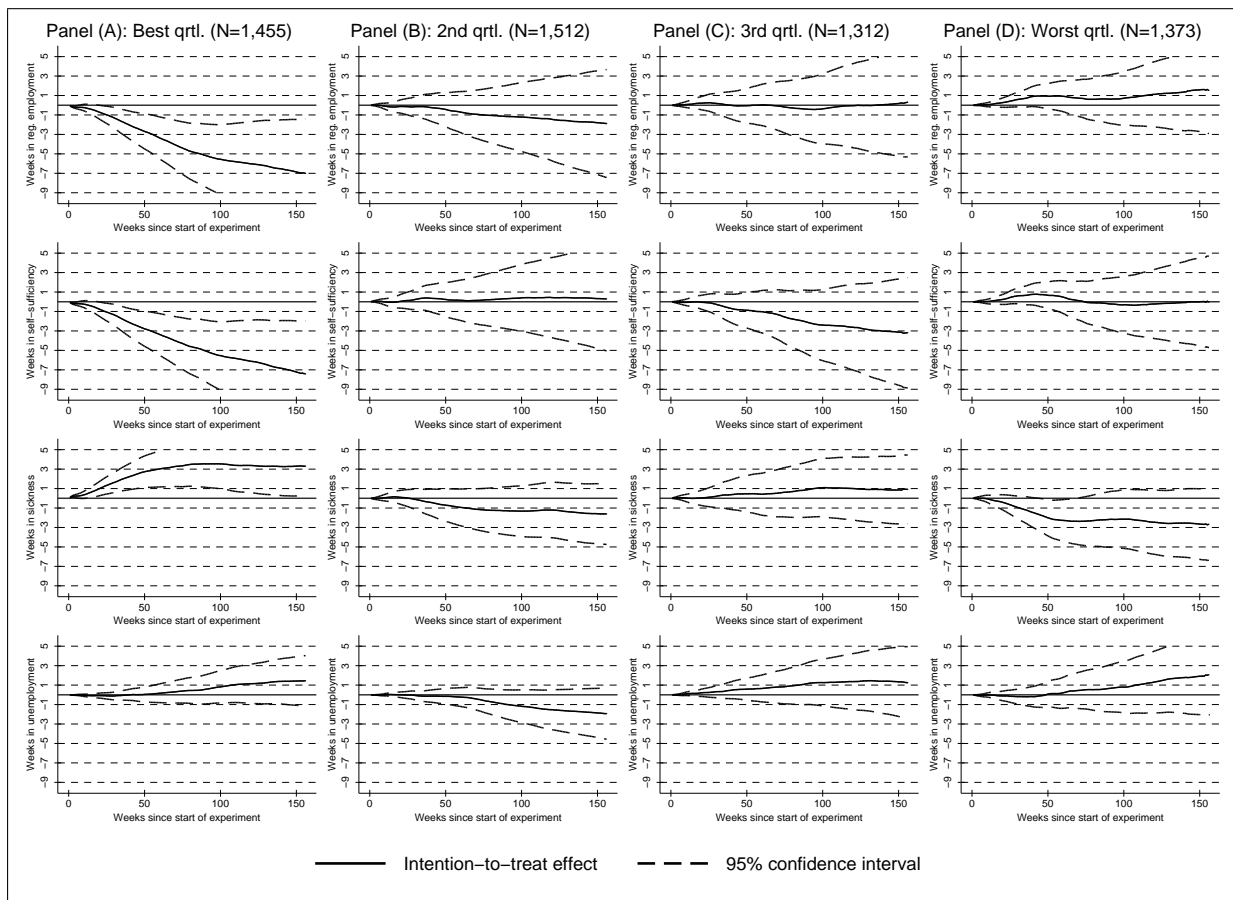
Notes: Indices 1 to 16 on the horizontal axes refer to job centers: (1) Bornholm, (2) Gentofte, (3) Greve, (4) København, (5) Ringsted, (6) Vordingborg, (7) Aalborg, (8) Morsø, (9) Randers, (10) Holstebro, (11) Herning, (12) Horsens, (13) Svendborg, (14) Nyborg, (15) Odense and (16) Aabenraa. Traditional activation intensities are calculated as the average number of traditional activation weeks per individual in the first 20 weeks after enrolment. Similarly for paramedical care intensities. Graded return intensities are calculated as the proportion of individuals participating in a graded return-to-work program within the first 20 weeks.

Figure A2: Trajectory of treatment effects by labor market status before sickness (cumulative total intention-to-treat effects)



Notes: See notes to Figure 2.

Figure A3: Trajectory of treatment effects by degree of benefit dependency in the year prior to assignment (cumulative total intention-to-treat effects)



Notes: See notes to Figure 2.